10

Ray tracing in a loss-free stratified medium

10.1. Introduction

The term 'ray' was used in § 5.3 when discussing the field of a radio wave that travels out from a source of small dimensions at the origin of coordinates ξ, η, ζ in a homogeneous medium. The field was expressed as an integral (5.29) representing an angular spectrum of plane waves. The main contribution to the integral was from 'predominant' values of the components n_ξ, n_η of the refractive index vector $\boldsymbol{n}$ such that the phase of the integrand was stationary for small variations δn_ξ, δn_η. For any point ξ, η, ζ there were one or more predominant values of the refractive index vector $\boldsymbol{n}$, such that the line from the origin to ξ, η, ζ was normal to the refractive index surface. Each predominant $\boldsymbol{n}$ defines a progressive plane wave that travels through the whole of the homogeneous medium.

Instead of selecting a point ξ, η, ζ and finding the predominant $\boldsymbol{n}$s, let us choose a fixed $\boldsymbol{n}$ and find the locus of points ξ, η, ζ for which this $\boldsymbol{n}$ is predominant. For a homogeneous medium this locus is called the 'ray'. The arguments leading to (5.31) still apply and show that it is the straight line through the origin, in a direction normal to the refractive index surface at the point given by the chosen $\boldsymbol{n}$. It was shown in § 5.3 that for a loss-free medium the direction of the ray is the same as the direction of the energy flux as given by the time averaged Poynting vector $\boldsymbol{\Pi}_{\text{av}}$, for the progressive plane wave that has the chosen $\boldsymbol{n}$.

These ideas may now be extended to apply to a plane stratified medium such as the ionosphere. We shall now therefore use the coordinate system x, y, z of § 6.1, with the z axis perpendicular to the strata. As in § 6.2 it is useful to imagine that the ionosphere is replaced by a number of thin discrete strata in each of which the medium is homogeneous. For a given progressive plane wave in any one stratum, the direction of the ray can be found. When this wave crosses a boundary, the x and y direction cosines S_1 and S_2 of its wave normal do not change; this is Snell's law, see § 6.2. But its value of q changes, so that in the next stratum the direction of the ray is different. In each stratum the ray is normal to the refractive index surface, but this gives only its direction. To find the actual curved ray path, the ray must be traced step by step starting from its source. In the limit when the strata are indefinitely thin and numerous, this leads to the integrals (10.3) below. The tracing of rays in this way has many uses of which the principal ones are as follows.

(a) When the direction of a ray reaching a receiver is known, the associated predominant $\boldsymbol{n}$ can be found and this gives the direction of the wave normal and therefore of the arriving wave front.

(b) For a loss-free medium the rays have the direction of the energy flux. By studying the configuration of adjacent rays in a ray pencil, the energy flux at the receiver can be found. This is discussed in §§ 10.17–10.20, 14.8.

(c) For a plasma with a very small collision frequency, the effect of the collisions on the real part of the refractive index $\boldsymbol{n}$ can often be neglected. The rays are traced using only the real part of $\boldsymbol{n}$ and are unaffected by the collisions. The attenuation of the signal is then found by suitably integrating the imaginary part of $\boldsymbol{n}$ along the ray path; see § 10.15.

(d) When pulsed signals are used, the pulse travels along the ray with the group velocity $\mathscr{U}$. The time of travel is found by integrating $1/\mathscr{U}$ along the ray path. See §§ 10.3, 10.16, 14.7.

10.2. The ray path

The radio transmitter may be thought of as a source of small dimensions in free space, at the origin of coordinates x, y, z, and emitting waves of fixed angular frequency $\omega = kc$, whose wave fronts are approximately spherical. Its radiated field is expressed, as in § 6.2, by an integral representing an angular spectrum of plane waves. In the free space below the ionosphere any one of the field components is then given by

$$F(x, y, z, t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} A(S_1, S_2) \exp\{ik(ct - S_1 x - S_2 y - Cz)\} \mathrm{d}S_1 \, \mathrm{d}S_2 \quad (10.1)$$

where the function $A(S_1, S_2)$ depends on the kind of transmitting aerial and on the field component used. In many cases of importance A is a very slowly varying function, and for small ranges of S_1, S_2 it can be treated as a constant when compared with the exponential in (10.1). This is a much more rapidly varying function of S_1, S_2, especially when kx, ky, kz are large. An important example where $A(S_1, S_2)$ is not slowly varying is described in § 10.3.

The integrand of (10.1) represents a plane progressive wave that travels obliquely upwards and enters the ionosphere. Within the ionosphere, for each S_1, S_2, there are two waves that can travel obliquely upwards, and these have different polarisations and different values of q, and they are reflected at different levels. When the incident wave enters the ionosphere, therefore, it divides into two waves that are propagated independently, and their relative amplitudes depend on the state of polarisation of the incident wave. It is possible to choose this so that only one magnetoionic component is present in the ionosphere. For this purpose, at moderate or high magnetic latitudes, and for near vertical incidence, the incident wave must be nearly circularly polarised. In practice it is more usual for the incident wave to be linearly polarised, and it then gives the two magnetoionic components with roughly equal amplitudes. This division of the incident wave into two magnetoionic components is related to the phenomenon called 'limiting polarisation', and the theory is given in §§ 17.10–17.11.

When the transmitted signal is a pulse of radio waves, as in the ionosonde technique, the incident pulse splits into two pulses that travel over different paths with different group velocities. After reflection they reach the receiver at different times so that the received signal or echo is said to be 'split'. The phenomenon is sometimes called magnetoionic splitting.

We now study one of the two magnetoionic component waves in the ionosphere. Let q be its associated root of the Booker quartic equation (ch. 6). In the free space below the ionosphere, $q = C$. It is assumed that, for levels z in the ionosphere far enough below the reflection level, the W.K.B. solution is sufficiently accurate. Below the ionosphere this is simply the factor $\exp(-\mathrm{i}kCz)$ in (10.1). Within the ionosphere this is replaced by a multiple of the W.K.B. solution derived from (7.136). It consists, as explained in § 7.18, of a phase memory factor $\exp(-\mathrm{i}k\int_0^z q\,\mathrm{d}z)$ and a more slowly varying factor that depends on which field component is used for F in (10.1). Thus (10.1) is replaced, within the ionosphere, by

$$F(x,y,z,t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} B(S_1,S_2,z)\exp\left\{\mathrm{i}k\left(ct - S_1x - S_2y - \int_0^z q\,\mathrm{d}z\right)\right\}\mathrm{d}S_1\,\mathrm{d}S_2. \tag{10.2}$$

Here the slowly varying part of the W.K.B. solution has been combined with the factor A in (10.1) to give the new slowly varying factor B.

The formula (10.2) fails if z is near a level of reflection, but it can be used for an obliquely downgoing wave that has been reflected at some greater value of z. This is done by arranging that the range of the integral $\int q\,\mathrm{d}z$ extends from the ground up to the level of reflection (in general different for each component wave) and then down again to the receiving point, with q replaced by a new q for the downgoing wave. This technique was explained in § 7.19.

The double integral (10.2) is of the type (9.61) and may be evaluated by the method of double steepest descents, § 9.10. There is a double saddle point where the partial derivatives of the exponent with respect to S_1 and S_2 are both zero; compare (9.63). The exponent in (10.2) may be denoted by $\mathrm{i}\varphi$ where φ is the phase. Then these conditions are equivalent to saying that the phase must be stationary with respect to variations of S_1 and S_2 (compare § 5.3). The conditions are

$$-\partial\varphi/\partial S_1 \equiv x + \int_0^z (\partial q/\partial S_1)\mathrm{d}z = 0, \quad -\partial\varphi/\partial S_2 \equiv y + \int_0^z (\partial q/\partial S_2)\mathrm{d}z = 0. \tag{10.3}$$

For any given position (x, y, z) of the receiver these might be solved for S_1, S_2 and would give $S_1 = S_{10}$, $S_2 = S_{20}$ say. Since B is slowly varying it may be replaced by $B(S_{10}, S_{20}, z)$. It is important only for small ranges of S_1 and S_2 near S_{10} and S_{20} respectively. If B fell to zero outside this range, the signal at the receiver would not be appreciably affected.

The value of the integral (10.2) is now

$$F(x, y, z, t) \approx G(S_{10}, S_{20}, z)\exp\left\{ik\left(ct - S_{10}x - S_{20}y - \int_0^z q\mathrm{d}z\right)\right\} \tag{10.4}$$

where G is a slowly varying function of z, and q takes its value for $S_1 = S_{10}$, $S_2 = S_{20}$. This gives the predominant progressive wave near the point (x, y, z).

We now trace the path traversed by the energy arriving at the receiver. Since this comes only from component waves with S_1 and S_2 near S_{10} and S_{20} respectively, we must find a series of points for which the signal is a maximum when $S_1 = S_{10}$, $S_2 = S_{20}$. The locus of these points is called the 'ray path' or simply the 'ray'. For each point on it the signal is a maximum when the phase is stationary for variations of S_1 and S_2. Hence (10.3) are the equations of the ray, provided that in $\partial q/\partial S_1$, $\partial q/\partial S_2$ we put $S_1 = S_{10}$, $S_2 = S_{20}$ after differentiation.

If (10.3) are differentiated with respect to z, they give

$$\mathrm{d}x/\mathrm{d}z = -\partial q/\partial S_1, \quad \mathrm{d}y/\mathrm{d}z = -\partial q/\partial S_2 \quad \text{for } S_1 = S_{10},\ S_2 = S_{20} \tag{10.5}$$

and these give the direction of the ray at each point. Now S_{10}, S_{20}, q are the coordinates of the point on the refractive index surface for the predominant progressive wave. Thus (10.5) confirms that the ray path is normal to the refractive index surface.

In general it is not easy to solve (10.3) for S_{10} and S_{20} when x, y, z are given. The more usual procedure is to assume a pair of values S_{10}, S_{20} and trace the ray corresponding to them. Successive pairs are tried until a pair is found that gives a ray passing through a given receiving point.

10.3. Wave packets

On a given ray the signal is the resultant of waves whose original direction cosines S_1, S_2 are within narrow ranges near S_{10}, S_{20} respectively. Suppose now that the transmitter radiates only these waves. Then the function $B(S_1, S_2, z)$ in (10.2) is no longer nearly independent of S_1, S_2, but has a narrow maximum near $S_1 = S_{10}$,

$S_2 = S_{20}$. It is no longer a slowly varying function, and the stationary phase condition (10.3) must be replaced by

$$x + \int_0^z \frac{\partial q}{\partial S_1} \mathrm{d}z + \frac{\mathrm{i}}{k}\frac{\partial(\ln B)}{\partial S_1} = 0, \quad y + \int_0^z \frac{\partial q}{\partial S_2} \mathrm{d}z + \frac{\mathrm{i}}{k}\frac{\partial(\ln B)}{\partial S_2} = 0. \tag{10.6}$$

If in these we put $S_1 = S_{10}$, $S_2 = S_{20}$, they give the equation of the ray. But the ray must pass through the transmitter, which is at the origin. Hence

$$\frac{\partial(\ln B)}{\partial S_1} = \frac{\partial(\ln B)}{\partial S_2} = 0 \tag{10.7}$$

when $S_1 = S_{10}$, $S_2 = S_{20}$. Therefore, although B is not now slowly varying, this adds nothing to the stationary phase condition. The resulting signal leaving the transmitter is in a narrow pencil of radiation with its maximum in the direction (S_{10}, S_{20}). This pencil enters the ionosphere and then has a curved path which is the ray path.

Suppose, further, that the signal is a pulse like that described in § 5.8. Then it must contain a range of frequencies, with maximum amplitude near the centre of the range. Since $k = 2\pi f/c$ is proportional to the frequency f, the pulsed signal may be expressed by assuming that B is a function also of k, and the expression becomes

$$F(x, y, z, t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} B(k, S_1, S_2, z) \exp\left\{\mathrm{i}k\left(ct - S_1 x - S_2 y - \int_0^z q\,\mathrm{d}z\right)\right\} \mathrm{d}k\,\mathrm{d}S_1\,\mathrm{d}S_2 \tag{10.8}$$

where B has a narrow maximum near

$$k = k_0, \quad S_1 = S_{10}, \quad S_2 = S_{20} \tag{10.9}$$

and these will be called the 'predominant' values of k, S_1, S_2. This represents a signal confined to a narrow pencil and of short duration. It is therefore a small packet of waves, and at a given time t it has maximum amplitude at some point (x, y, z). As t increases, the locus of this point is the path of the wave packet. The wave packet is assumed to leave the transmitter when $t = 0$ and this requires that $\partial(\ln B)/\partial k = 0$ when (10.9) is satisfied.

The signal (10.2) is a maximum when the phase is stationary with respect to variations of k, S_1, S_2 at the predominant values (10.9). This gives the two equations (10.3), and in addition

$$ct - S_1 x - S_2 y - \int_0^z \frac{\partial(kq)}{\partial k} \mathrm{d}z = 0. \tag{10.10}$$

Here t is the time taken for the wave packet to travel from the origin to the point (x, y, z) where x and y are given in terms of z by (10.3). In this time the wave packet could travel a distance $P' = ct$ in free space, and this is called the 'equivalent path'.

Thus (10.10) can be rewritten

$$P' = S_1 x + S_2 y + \int_0^z \frac{\partial(fq)}{\partial f} \mathrm{d}z \tag{10.11}$$

where (10.9) is satisfied.

Although most transmitters do not radiate wave packets, we have shown that the problem of finding the path of the energy reaching a given receiver is the same as that of finding the path of a wave packet. In future, therefore, we shall speak of a ray and of the path of a wave packet as the same thing.

10.4. Equations of the ray path

It has been shown that for a ray path in a plane stratified medium, S_1 and S_2 have constant values S_{10} and S_{20}. In the discussion of rays that follows, the second subscript 0 will be omitted, and we use simply S_1, S_2 for the constant values on a ray.

The coordinates x, y, z of a point on a ray in a stratified medium are related by the two differential equations (10.5), or by the two equations in integrated form, (10.3). These were first given by Booker (1939) and are now widely known as Booker's ray tracing equations.

Assume first that the ionospheric plasma is isotropic. Then q is given, from (6.6), by

$$q^2 = n^2 - S_1^2 - S_2^2 \tag{10.12}$$

where n is the refractive index and is independent of S_1 and S_2. It was shown in § 10.3 that a ray is the same as the path of a wave packet and that on it S_1 and S_2 are constant. Hence (10.5) give

$$\mathrm{d}x/\mathrm{d}z = S_1/q, \quad \mathrm{d}y/\mathrm{d}z = S_2/q \tag{10.13}$$

and

$$x = S_1 \int_0^z \frac{\mathrm{d}z}{q}, \quad y = S_2 \int_0^z \frac{\mathrm{d}z}{q}. \tag{10.14}$$

These show that on a ray the ratio x/y is a constant so that the ray remains in the same vertical plane. It is nearly always convenient to choose the axes so that this is the plane $y = 0$, and $S_2 = 0$.

To find $\partial q/\partial S_1$ and $\partial q/\partial S_2$ for the general anisotropic plasma we make this choice of axes from the start, so that $S_2 = 0$, $S_1 = S$. The Booker quartic (6.15) must be true for all values of S_1 and S_2. Hence

$$\mathrm{d}F(q)/\mathrm{d}S_1 = \mathrm{d}F(q)/\mathrm{d}S_2 = 0. \tag{10.15}$$

Now

$$\frac{\mathrm{d}F(q)}{\mathrm{d}S_1} = \frac{\partial F(q)}{\partial q}\frac{\partial q}{\partial S_1} + \frac{\partial \alpha}{\partial S_1}q^4 + \frac{\partial \beta}{\partial S_1}q^3 + \frac{\partial \gamma}{\partial S_1}q^2 + \frac{\partial \delta}{\partial S_1}q + \frac{\partial \varepsilon}{\partial S_1} = 0 \tag{10.16}$$

and there is a similar equation with S_1 replaced by S_2. Equation (6.16) shows that $\partial\alpha/\partial S_1 = \partial\alpha/\partial S_2 = 0$. Hence, for a ray path, from (10.5)

$$\left.\begin{aligned}\frac{\mathrm{d}x}{\mathrm{d}z} &= \left(\frac{\partial\beta}{\partial S_1}q^3 + \frac{\partial\gamma}{\partial S_1}q^2 + \frac{\partial\delta}{\partial S_1}q + \frac{\partial\varepsilon}{\partial S_1}\right)\Big/\frac{\partial F(q)}{\partial q}\\ \frac{\mathrm{d}y}{\mathrm{d}z} &= \left(\frac{\partial\beta}{\partial S_2}q^3 + \frac{\partial\gamma}{\partial S_2}q^2 + \frac{\partial\delta}{\partial S_2}q + \frac{\partial\varepsilon}{\partial S_2}\right)\Big/\frac{\partial F(q)}{\partial q}\end{aligned}\right\}\begin{aligned}&\text{when } S_1 = S\\ &\text{and } S_2 = 0.\end{aligned} \tag{10.17}$$

Curves have been given by Booker (1949) to show how these quantities depend on X in some special cases.

The second expression in (10.17) is not in general zero even when $S_2 = 0$. Thus a ray that starts out in the vertical plane $y = 0$ does not necessarily remain in that plane. It may be laterally deviated. Some examples of this lateral deviation are given in §§ 10.12–10.14.

In some radio problems it is necessary to allow for the earth's curvature. Consider therefore a spherically stratified isotropic ionosphere, so that the refractive index $n(r)$ is a function of distance r from the centre of the earth. At radius r let the wave normal of a progressive wave make an angle ψ with the radius. Because the medium is now assumed to be isotropic, this is also the direction of the ray. Now ψ satisfies Bouger's law

$$rn\sin\psi = K \tag{10.18}$$

where K is a constant. This replaces Snell's law (6.4) for a plane stratified medium. Equation (10.18) is easy to prove by simple geometry (see problem 10.3). If r, θ are the polar coordinates of any point on a ray path

$$\frac{r\mathrm{d}\theta}{\mathrm{d}r} = \tan\psi = \frac{K}{rn\cos\psi} = \frac{K/r}{q} \tag{10.19}$$

where

$$q^2 = n^2 - K^2/r^2. \tag{10.20}$$

Equation (10.19) is the differential equation of a ray path in various forms. It should be compared with (10.13). An observer who thought that the earth was flat would treat dr as an element of height dz, and $r\mathrm{d}\theta$ as an element of horizontal distance dx. Then (10.19), (10.20) are the same as the first equation (10.13), and (10.12) with $S_2 = 0$, provided that we take

$$S_1 = K/r. \tag{10.21}$$

This shows that, for the curved earth, the ionosphere can be treated as though it is plane stratified provided that we use (10.21), so that S_1 is no longer constant on a ray. This device is called an 'earth flattening method'. In the present example, with an isotropic ionosphere, it is exact. For an anisotropic ionosphere it can also be used but then involves approximations, and we speak of an 'earth flattening approximation'. This topic is discussed further in §18.8.

Equations (10.19), (10.20) can also be written

$$r\mathrm{d}\theta/\mathrm{d}r = S_0/\bar{q}, \quad \bar{q}^2 = \bar{n}^2 - S_0^2, \quad \bar{n} = nr/r_0 \tag{10.22}$$

where r_0 is some fixed value of r, for example at the earth's surface, and S_0 is a constant equal to the value of $n \sin\psi$ where $r = r_0$. Thus (10.22) provides an alternative 'earth flattening' method, in which S_0 is constant on a ray. Instead of n it uses the modified value $\bar{n}$. This is very similar to the modified refractive index used by Booker and Walkinshaw (1946) in one of the pioneering papers on guided radio waves. For a further study of equations (10.18)–(10.20) see § 10.23.

10.5. The reversibility of the path

A wave packet leaves the transmitter, is deviated by the ionosphere and returns to the ground. Suppose now that every component plane wave in the downcoming wave packet is reversed in direction. This means that both S_1 and S_2 are reversed in sign, and (6.16) shows that the only effect on the quartic is to reverse the signs of the coefficients of q and q^3. Thus q also changes sign, so that $\partial q/\partial S_1$ and $\partial q/\partial S_2$ remain unchanged and (10.5) shows that the direction of the path is the same as before. The wave packet therefore simply retraces its original path back to the transmitter. The state of polarisation of the wave in a wave packet can be found from (6.21), which gives the ratios of the three components E_x, E_y, E_z. These depend on the elements of the 3×3 matrix in (6.21), (6.22), which remain unaltered when q, S_1 and S_2 all change sign. Hence the polarisation referred to fixed axes x, y, z is the same for the original and the reversed wave packet. This has an important bearing on the study of reciprocity; §§ 14.13, 14.14.

The equations (10.5) were based on the assumption that the ray is a feature of a wave system of the form (10.2) in which the z dependence of the exponent is given entirely by the factor $\int_0^z q\,\mathrm{d}z$. Here q refers to one of the four magnetoionic components and it is implied that this is propagating independently of the other three. A wave can, however, give rise as it travels to some of the other components, and these react back on the original wave and modify its properties. The subject is studied in ch. 16 on coupled wave equations, where it is shown, § 16.9, that q must be replaced by $q + \mathrm{i}\Gamma_{jj}$. In many practical cases the extra term is negligible. In cases where Γ_{jj} cannot be neglected, the ray tracing equations (10.5) must be modified. It is then found that the rays are not reversible. This subject has been studied by Smith, M.S. (1975, 1976) who gives some examples.

10.6. The reflection of a wave packet

In an isotropic ionosphere a wave is reflected at the level where the ray and the wave normal are horizontal, that is where $q = 0$. Here the integrands in (10.14) are infinite. These integrals, however, always converge to a finite limit when z is at the reflection level. To illustrate this consider a simple example where the electron height distribution is linear so that $X = pz$. Let $S_2 = 0$, $S_1 = S$. Then

$$q^2 = n^2 - S^2 = C^2 - X = C^2 - pz. \tag{10.23}$$

Let a ray leave the origin in this medium. To evaluate the integral (10.14) it is often simplest to change to q as the variable of integration. This gives

$$x = 2S \int_C^q \frac{\mathrm{d}q}{\mathrm{d}(q^2)/\mathrm{d}z} = \frac{2S}{p}(C - q). \tag{10.24}$$

Clearly this converges to the limit $2SC/p$ at the reflection level where

$$q = 0, \quad z = z_0 = C^2/p. \tag{10.25}$$

The ray path (10.24) is the parabola

$$S^2 z = SCx - \tfrac{1}{4}px^2. \tag{10.26}$$

It was shown in §§7.19, 8.20 that when an upgoing wave represented by the integrand of (10.2) reaches the level of reflection $z = z_0$, it is converted to a downgoing wave, whose fields are given simply by using the factor (7.155) in the integrand. This gives the phase memory for the ray path up to the reflection level and then down again after reflection. Thus the ray can be traced right through the reflection level as a continuous curve. The phase advance factor i mentioned at the end of § 8.20 does not affect the ray path. In the present example the parabolic path (10.26) extends beyond its maximum to the descending part.

For an anisotropic ionosphere at a reflection level, two roots of the Booker quartic, say q_1 and q_2, are equal. The denominators of the ray equations (10.17) are $\partial F(q)/\partial q$ which has a factor $q_1 - q_2$. Thus integration of (10.17) uses an integral with an infinite integrand at the reflection level. But again it can be shown that the integral must converge to a limit. This is because $(q_1 - q_2)^2$ is an analytic function of z at $z = z_0$, like q^2 in (10.23). For the proof see § 16.3. Thus in an anisotropic medium also, a ray can be traced right through a reflection level.

It is possible for the quartic to have two equal roots which differ from zero. Where this happens the path of the ray is horizontal, but the direction of the wave normal is not horizontal. The wave packet does not in general travel in the direction of the wave normal. Similarly, at a level where one value of q is zero, the associated wave packet is not necessarily travelling horizontally, although the wave normal is horizontal. Thus the level where $q = 0$ is not a level of reflection, except in the special case where there are two roots which are both zero at the same level.

The first expression (10.17) is $\mathrm{d}x/\mathrm{d}z$ for the ray path. It can change sign at a level where the numerator is zero and the path of the wave packet is then parallel to the y–z plane, and the x component of its motion is reversed. Some examples of this can be seen in fig. 10.10. The sign of the vertical component of its motion is unaltered, however, and can only be reversed where $\partial F(q)/\partial q$ is zero.

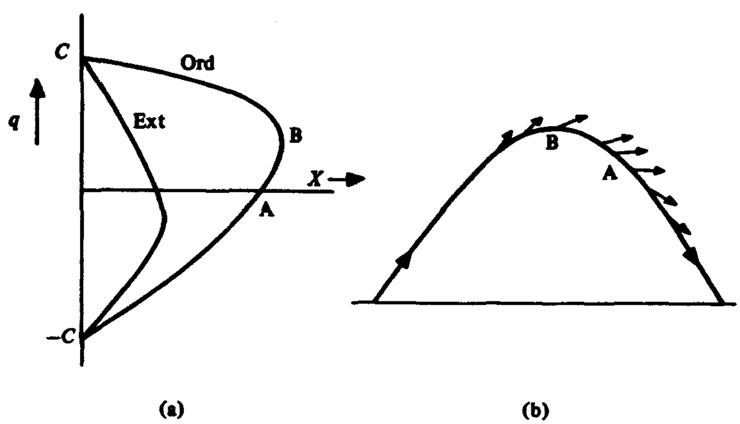
In § 6.4 it was shown that when S_1, S_2 are real two of the four waves are obliquely upgoing waves and two are obliquely downgoing.

10.7. An example of a ray path at oblique incidence

Suppose that the frequency is greater than the gyro-frequency, and consider a ray that leaves the earth obliquely in the magnetic meridian plane. It is shown later, § 10.14, that there is no lateral deviation in this case, so the ray remains in the magnetic meridian throughout its path. We take $S_2 = 0$ so that the x–z plane is the plane of incidence and also the magnetic meridian plane. Fig. 10.1(a) shows an example of how the four roots of the quartic depend on X in this case. When the ray enters the ionosphere it splits into ordinary and extraordinary rays as described in § 10.2. The direction of the ordinary ray is given at each level by (10.17), first equation, and the ray can thus be plotted as shown in fig. 10.1(b). When the level B is reached, two roots of the quartic are equal and the ray is horizontal, but q is not zero and the wave normal is still directed obliquely upwards because q is positive. When the downgoing ray reaches the level where $q = 0$, at A in figs. 10.1(a, b), the wave normal is horizontal. Thereafter, on the rest of the downward path, the wave normal is obliquely downwards. The behaviour of the extraordinary ray is similar, but the level of reflection is lower than for the ordinary ray, and the wave normal becomes horizontal before reflection occurs.

Another example of a ray path is shown in fig. 10.2. It is for a frequency less than the electron gyro-frequency, $Y = 2$, and again is for a ray in the magnetic meridian plane. The dependence of q on X is shown by the curve marked Ext in fig. 6.7. There are vertical tangents at the three points D, A, B and the ray path is horizontal at the

Fig. 10.1. Finding ray paths at oblique incidence. In this example the incident ray is in the magnetic meridian plane and travelling southwards. (a) shows how the four roots q of the Booker quartic depend on X. (b) shows the path of the ordinary ray and the arrows show the direction of the wave normal at each level.



three corresponding levels, as shown in fig. 10.2. The wave normal is horizontal where $q = 0$ at a level just below B, as shown by the arrows.

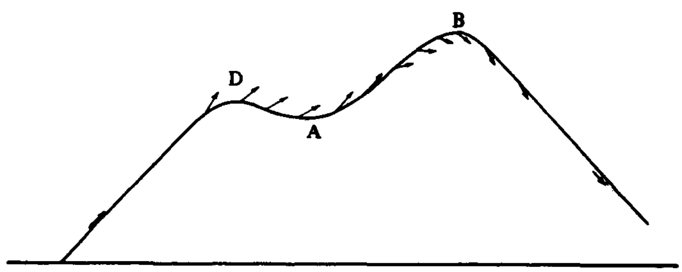
10.8. Poeverlein's construction

A method for finding the general shape of a ray path in a stratified ionosphere was given by Poeverlein (1948, 1949, 1950). In the free space below the ionosphere the incident ray is in the x–z plane at an angle θ to the z axis. It enters the ionosphere and there splits into two rays, ordinary and extraordinary. We study one of these, say the ordinary ray. A diagram is drawn with axes n_x, n_z parallel to the x, z axes of ordinary space. It is thus a cross section of refractive index space by a plane parallel to the plane of incidence. A sequence of values of X is now selected, and for each X the cross section of the refractive index surface for the ordinary wave is drawn in the diagram. An example is shown in fig. 10.3. In general the earth's magnetic field is not in the plane of this diagram, but its projection is shown in the figure. The outermost curve is for the free space below the ionosphere and is a circle of unit radius. As the electron concentration increases the curves get smaller until they shrink to a point at the origin when $X = 1$.

Now let a vertical line, AB in fig. 10.3, be drawn in the $n_x - n_z$ plane at a distance $S = \sin\theta$ from the origin; compare fig. 6.1. It is here called the reference line. For some level let it cut the refractive index surface at the two points C and D, and let OC make an angle ψ with the n_z axis. Then OC is the refractive index n for a wave whose normal is in the direction OC, and $\mathrm{OC}\sin\psi = n\sin\psi = S$, which is simply Snell's law. Hence OC is one possible direction for the wave normal of the ordinary wave. Another possible direction is OD, and two other directions are found from the intersection of the line AB with the refractive index surface for the extraordinary wave. It is also clear from the figure that $\mathrm{CE} = n\cos\psi = q$.

Now it was shown in § 5.3 that the ray direction is perpendicular to the refractive index surface. Hence the two possible ray directions for the ordinary wave are the perpendiculars to the surface at C and D. Clearly the one at C must be inclined

Fig. 10.2. An example of a ray path in the magnetic meridian plane. The dependence of q on X for this example is shown by the curve marked Ext in fig. 6.7.



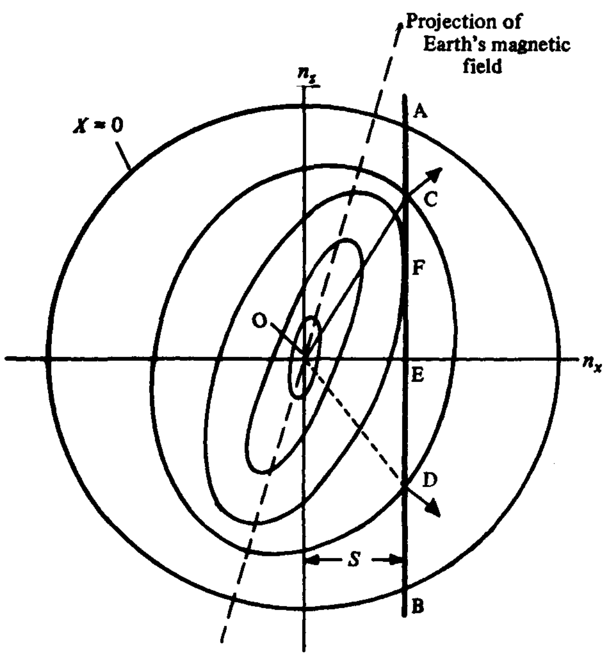
upwards and that at D downwards. These ray directions are not in general in the plane of the figure, that is the n_x–n_z plane. For example, that at C must be coplanar with OC and the earth's magnetic field.

The reference line cuts the circle of unit radius at A so that OA is the direction in which the ray enters the ionosphere. As the wave packet travels upwards, the refractive index surface changes, so that the direction of the wave normal and the ray also change. The wave normal is always in the plane of the diagram, but the ray may not be. The points C and D move closer together until a level is reached where the reference line just touches a refractive index surface at F in fig. 10.3. Then the ray direction is horizontal and the two associated values of q are equal. This is the level of reflection, and thereafter the wave packet travels downwards and we consider the same series of surfaces as before, but in reverse order. Finally the surface of unit radius is again reached, and the wave packet leaves the ionosphere with the ray and wave normal both in the direction OB.

It is clear that in general the ray is directed out of the plane of the diagram. The only exception to this is for propagation from (magnetic) north to south or south to north, and this case is discussed separately in the following sections.

If Poeverlein's construction is applied to a real ray that originates in the free space

Fig. 10.3. To illustrate Poeverlein's construction. Cross section by the plane of incidence of a sequence of refractive index surfaces for the ordinary ray, for various values of X.

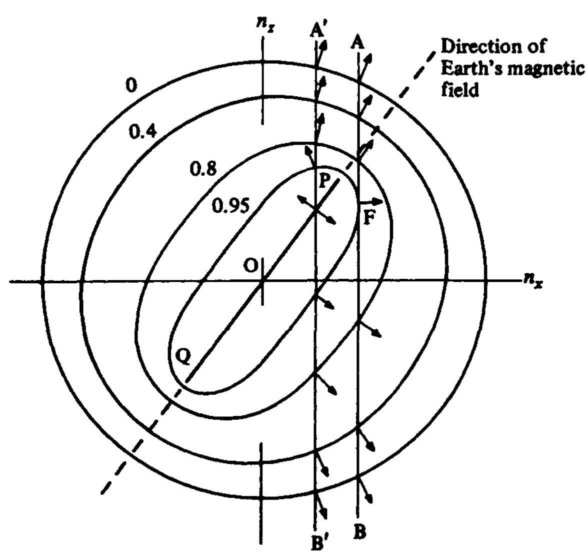


below the ionosphere, the angle of incidence θ is real so that S must be real and $|S| < 1$. But in a plasma the refractive index can exceed unity and then fig. 10.3 might contain refractive index surfaces where $|n| > 1$. Some of these could be real surfaces in the region where $|S| > 1$. There are examples in figs. 10.7, 10.9. A reference line AB in fig. 10.3 with $S > 1$ would cut such a surface at real points and here the wave normal and the ray would have real directions. Such a ray could be propagated in a horizontally stratified ionosphere, although it could not originate from a ray with real θ in free space. It might come from a transmitter in a space vehicle within the ionosphere, or from outside the ionosphere by refraction where there are horizontal variations. Signals of this kind have been observed. An example is the subprotonic whistler described at the end of § 13.8.

10.9. Propagation in magnetic meridian plane. The 'Spitze'

Fig. 10.3 shows a series of cross sections of the refractive index surfaces by the plane of incidence, that is the n_x–n_z plane. When this plane is also the magnetic meridian, these cross sections, for the ordinary ray, have a slightly different form, and an example is shown in fig. 10.4. As the electron concentration increases, the curves get smaller, as in fig. 10.3, but instead of shrinking to a point when $X = 1$, they shrink to the line PQ. This shows that the refractive index n for the ordinary wave is zero when $X = 1$ for all directions except the direction of the earth's magnetic field. For this case

Fig. 10.4. Similar to fig. 10.3 but the plane of incidence is now the magnetic meridian. In this example $Y = \frac{1}{2}$ (compare fig. 5.4, left-hand figure). The numbers by the curves are the values of X.

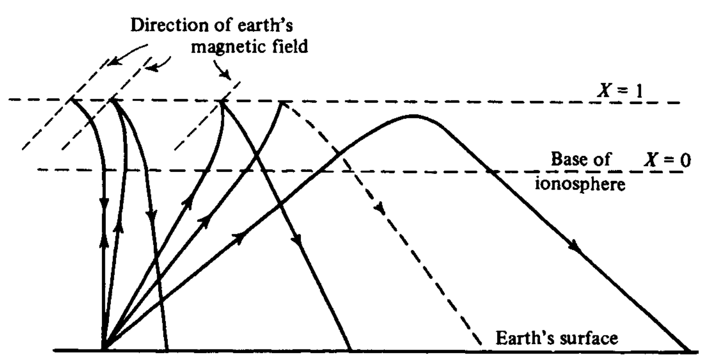


of purely longitudinal propagation there is an ambiguity in the value of n when $X=1$, mentioned in §4.11.

Poeverlein's construction may now be used as described in the previous section. A vertical reference line is drawn at a distance S from the origin. Suppose it is at AB in fig. 10.4. Then the ray directions are given by the perpendiculars to the refractive index surfaces where the line cuts them. They are now always in the plane of incidence, so there is no lateral deviation. Suppose that a wave packet travels upwards in an ionosphere in which the electron concentration increases monotonically. Eventually a level is reached where the reference line just touches a refractive index surface, as at F in fig. 10.4. Here the ray is horizontal and this is the level of reflection. The phenomenon is similar to that described in the previous section. The surface that the line just touches gives the value of X at the level of reflection. If the angle of incidence, arcsin S, increases, the reference line moves further from the origin and the surface which is touched corresponds to a smaller value of X. The reflection level therefore decreases as the angle of incidence increases.

Suppose, however, that the reference line is at A′B′ in fig. 10.4. The ray directions are given as before by the perpendiculars to the refractive index surfaces where the line cuts them, but there is now no surface which is touched by the line. Instead, the line A′B′ cuts the line PQ, which refers to the level where $X=1$. Above PQ the outward perpendicular to the refractive index surface is directed obliquely upwards; below PQ the perpendicular is obliquely downwards. On the line PQ the ray therefore reverses its direction abruptly, and this is the level of reflection. Here the ray path is perpendicular to the line PQ, that is to the earth's magnetic field. The ray path never becomes horizontal, but has a cusp at the level $X=1$, called by Poeverlein the 'Spitze'. Fig. 10.5 shows some typical ray paths for various angles of incidence. It is similar to a diagram given by Poeverlein (1950). As long as the angle

Fig. 10.5. Typical ray paths for the ordinary ray in the magnetic meridian, showing the Spitze.



of incidence is such that the line A′B′ cuts the line PQ, there is always a Spitze, and the reflection level is where $X = 1$ for all angles of incidence.

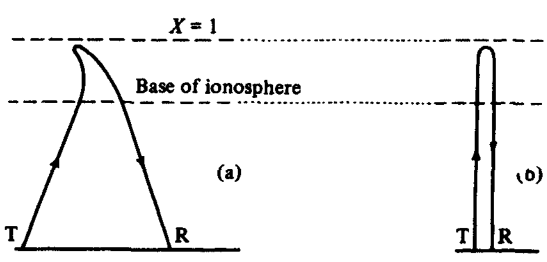
If the reference line is drawn where S is given by (6.32), plus sign, it goes through P. Then the upgoing ray reaches the level where $X = 1$ but is not now reflected. It is converted to an extraordinary ray in the Z-mode, above the level where $X = 1$, and its path is described in § 10.10, fig. 10.8(a). The incident ray is in the magnetic meridian at an angle θ_w to the vertical. For directions near this in a small cone whose width is a few degrees, the upgoing ray is partially reflected and partially converted to a Z-mode ray. This small cone of angles is known as a 'radio window'. The phenomenon cannot be explained by simple ray theory alone and a full wave treatment is needed, given in §§ 17.6–17.9. There is a similar effect if the reference line in fig. 10.4 goes through Q; see fig. 10.8(b). For this reason the points P and Q in refractive index space are known as window points.

If the plane of incidence is turned very slightly from the magnetic meridian, the line PQ moves out of the plane of the diagram in fig. 10.4. The line A′B′ then touches one of the refractive index surfaces, for X slightly less than unity. This is a long cigar-shaped surface close to PQ. The perpendicular to it is a horizontal line almost perpendicular to the plane of the diagram. The ray path is then a twisted curve in three dimensions, and fig. 10.6(a, b) shows how it would appear when seen from the west and the north respectively. The receiver is not in the plane of incidence through the transmitter, for the reasons explained in § 10.14. A ray path with a Spitze is a limiting case, when the width of the curve of fig. 10.6(b) in the east–west direction becomes infinitesimally small.

10.10. Ray paths for the extraordinary ray when $Y < 1$

When $Y < 1$ so that the frequency is greater than the gyro-frequency, one set of refractive index surfaces, for the extraordinary ray, extends from $X = 0$ to $X = 1 - Y$; see fig. 5.7 left-hand diagram. The outermost curve for $X = 0$ is a circle of unit radius. The curves get smaller as X increases and eventually shrink to a point

Fig. 10.6. Projections of the ray path for the ordinary ray when the plane of incidence is slightly inclined to the magnetic meridian. (a) is projected on to the plane of incidence and (b) on to a vertical plane at right angles to it.

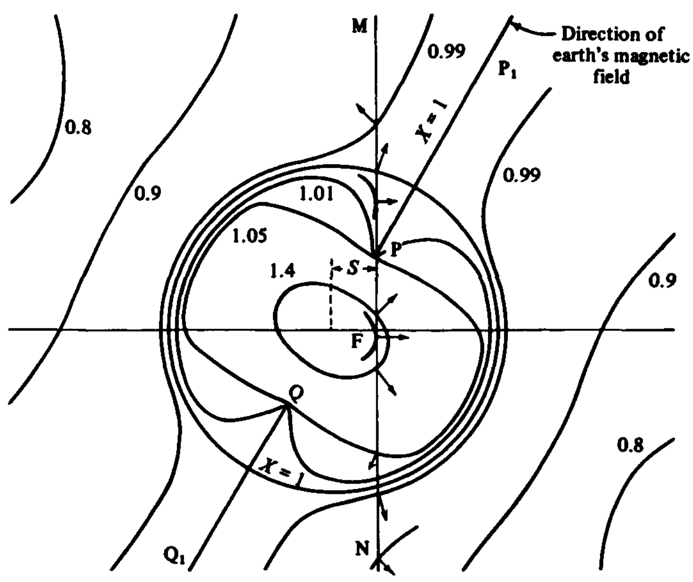


when $X = 1 - Y$. They never shrink to a line as in fig. 10.4 so that there is no phenomenon of the Spitze for the extraordinary ray when $X < 1$. The value of X at reflexion decreases as S increases, for the whole range $0 \leqslant S \leqslant 1$.

The refractive index for the extraordinary ray is imaginary when $1 - Y < X < 1 - Y^2$, but for $X > 1 - Y^2$ it is again real, and another set of refractive index surfaces can be constructed. The cross section of these by the magnetic meridian is shown in fig. 10.7. Now some values of n are greater than unity, so that there are some curves outside the unit circle. There is no curve for $X = 0$. The curves shrink to zero when $X = 1 + Y$. When $X = 1$ (see figs. 5.9, 5.11), the curve includes the unit circle, but it also includes the two straight line segments PP_1 and QQ_1 which are along the direction of the earth's magnetic field. This is because for purely longitudinal propagation there is an ambiguity in the value of n when $X = 1$. The points P and Q in fig. 10.7 are the window points and are the same as those in fig. 10.4. The two sets of surfaces have these points in common.

In general, for a wave packet incident from below, the extraordinary ray could not reach the levels depicted in fig. 10.7, for it would be reflected at or below the level where $X = 1 - Y$. There is a special case, however, when the angle of incidence is such that the reference line in fig. 10.4 goes through the point P. This figure refers to the ordinary wave, but at the level where $X = 1$ the wave is converted into an extraordinary wave, and the remainder of its path is found from fig. 10.7. The line AB

Fig. 10.7. Cross section, by the magnetic meridian plane, of the second set of refractive index surfaces for the extraordinary ray when $Y = \frac{1}{2}$. Compare figs. 5.8, 5.9. The numbers by the curves are the values of X.

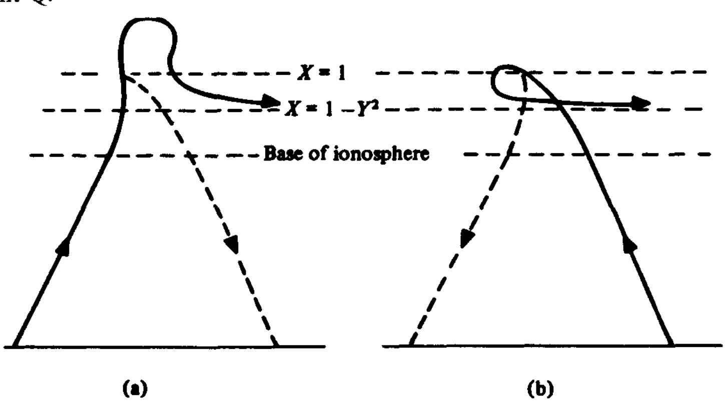


of fig. 10.4 must then be continued in fig. 10.7 and is there shown as the line MN. The two parts PM and PN correspond to different rays. The part PM near to P gives a ray travelling downwards and arriving at the level $X = 1$ from above, and so this ray cannot be present. The part PN gives a ray in the Z-mode that travels up to the level where the line PN just touches a refractive index surface at F. Here the ray is horizontal so this is a level of reflection. The ray then comes down again and crosses the level where $X = 1$, but can never leave the ionosphere because the level $X = 0$ does not appear in fig. 10.7. The line PN successively cuts curves of increasing radius and the ray becomes more and more horizontal. The wave packet therefore travels nearly horizontally just above the level where $X = 1 - Y^2$. In fact its energy would ultimately be absorbed because of electron collisions, which cannot be neglected in this case. The ray path is shown in fig. 10.8(a). The behaviour is similar when the line in figs. 10.4 and 10.7 goes through the point Q. The ray path for this case is shown in fig. 10.8(b). For a more detailed study of these ray paths see Poeverlein (1950, p. 157).

10.11. Extraordinary ray when $Y > 1$

When $Y > 1$, the cross sections by the magnetic meridian plane of the refractive index surfaces for the extraordinary ray are as shown in fig. 10.9 which is a combination of the left-hand diagrams of figs. 5.10–5.13. As X approaches $1 + Y$, the curves get smaller and eventually shrink to a point. The unit circle is the curve for $X = 0$. The curve for $X = 1$ is also the unit circle, but in addition it includes the two straight line segments PP_2, QQ_2 which are along the direction of the earth's magnetic field. The four points P, Q, P_2, Q_2 are window points, and P, Q are the same as in fig. 10.4. For

Fig. 10.8. Ray path for an incident ordinary ray in the magnetic meridian plane, at the transitional angle of incidence. It is converted to an extraordinary ray in the Z-mode at a window point. In (a) the lines AB in fig. 10.4 and MN in fig. 10.7 pass through the window point P. In (b) these lines pass through the window point Q.



$X < 1$ the refractive index curves have an envelope shown in the inset of fig. 10.9 (see problem 5.9).

When the frequency is small enough to make $Y > 1$, it is not usually permissible to neglect electron collisions, so that the curves of fig. 10.9, in which collisions are neglected, cannot be applied to the actual ionosphere. It is, nevertheless, instructive to use fig. 10.9 to plot ray paths in the magnetic meridian, for a fictitious ionosphere in which collisions are absent. This is done by Poeverlein's construction exactly as in §§ 10.8–10.10. A vertical reference line AB is drawn at a distance S from the origin. Six examples are shown in fig. 10.10. For the line A_1B_1 (fig. 10.9) the ray path is as in fig. 10.10(a), and for the line A_2B_2 it is as in fig. 10.10(b). In both these cases when the upgoing ray passes the level where $X = 1$, its direction is parallel to the incident ray, and when the downgoing ray passes this level its direction is parallel to the emergent ray below the ionosphere. Fig. 10.10(c) shows the ray path for the line A_3B_3 in fig. 10.9. This crosses the straight segment PP_2 to the right of the point R and to the left of the right branch of the envelope, so that it does not intersect this branch. When the ray first reaches the level where $X = 1$, the corresponding point is S in

Fig. 10.9. Cross section, by the magnetic meridian plane, of the refractive index surfaces for the extraordinary ray when $Y = 2$. The numbers by the curves are the values of X. Inset is an enlarged view of part of the diagram.

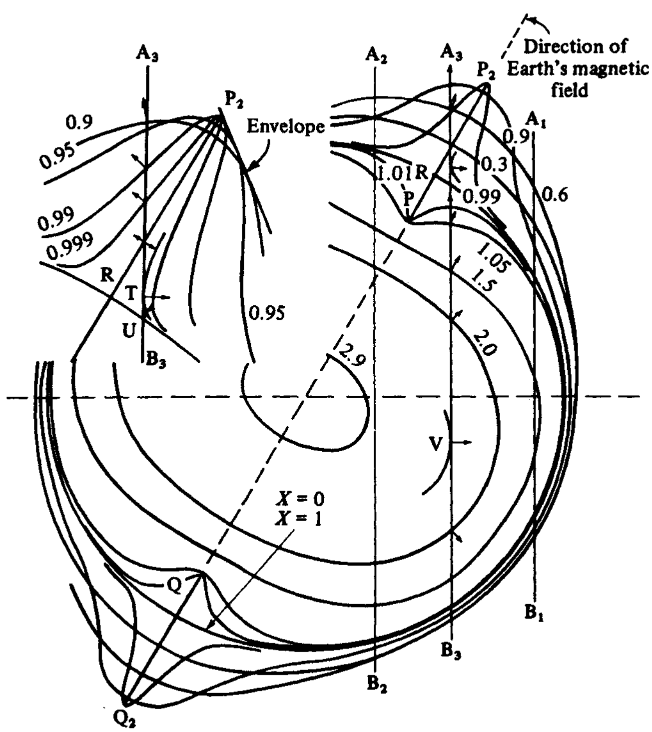
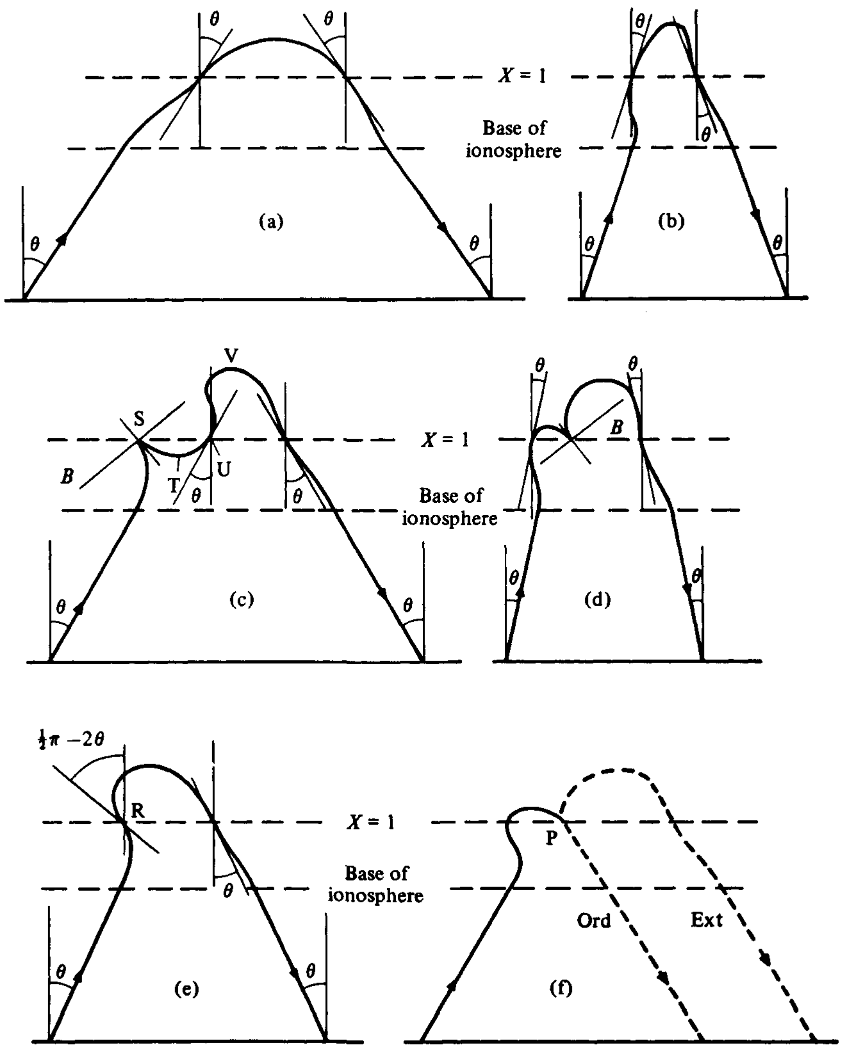


fig. 10.10(c). The ray direction is here suddenly reversed, in exactly the way described in § 10.8 for the ordinary ray, and there is a 'Spitze' in the ray path. The ray next goes down to a level where it is horizontal, corresponding to the point T in the figures. Then it rises again, passes through the level where $X = 1$ (point U, figs. 10.9 and

Fig. 10.10. Ray paths for the extraordinary ray in the magnetic meridian plane when $Y > 1$ and electron collisions are neglected. In each diagram the angle marked θ is equal to the angle of incidence. In (c) and (d) the line B shows the direction of the earth's magnetic field. The position of the line AB in fig. 10.9 is as follows. For (a) at A_1 B_1, for (b) at A_2 B_2, and for (c) at A_3 B_3. For (d) it is between P and R, for (e) it passes through R and for (f) it passes through the window point P.



10.10(c)), and again becomes horizontal at the level corresponding to point V. Thereafter it travels downwards, again passing the level where $X = 1$, and eventually reaches the ground.

Fig. 10.10(d) shows the ray path when A_3B_3 passes between P and R in fig. 10.9, and fig. 10.10(e) is the path when A_3B_3 passes through R. The detailed tracing of these rays may be left as an exercise for the reader, who should verify that at the level where $X = 1$, corresponding to the point R, the ray is inclined to the vertical at an angle $\frac{1}{2}\pi - 2\theta$ where θ is the angle of incidence.

When the line AB in fig. 10.9 passes through the point P the ray path can be traced as before until the representative point reaches P. The extraordinary wave can here be converted into an ordinary wave, for the point P is common to the surfaces in fig. 10.9 and 10.4. The rest of the ray path is then found from fig. 10.4, and the ray returns to earth as an ordinary wave although it started out as an extraordinary wave. The complete ray path in this case is as shown in fig. 10.10(f). The wave would split into an ordinary and extraordinary wave for a cone of angles of incidence near the critical value, as explained in § 10.9.

The case where the reference line goes through the window point P_2 (or Q_2) is not given in detail here. This point is common to fig. 10.9 and to the set of refractive index surfaces for the ordinary wave when $Y > 1$, $X > 1$, that is for the whistler mode, fig. 5.5 left diagram. The points P_2, Q_2 are window points sometimes referred to as the 'second window' (Jones, 1976b; Budden, 1980); § 17.9.

10.12. Lateral deviation at vertical incidence

The Booker quartic may be used to find the path of a wave packet that is vertically incident on the ionosphere from below. It is convenient to choose the x axis to be in the magnetic meridian, pointing north, so that the direction cosine $l_y = 0$. The partial derivatives with respect to S_2 of all five coefficients α to ε in (6.16) are zero when $S_2 = 0$, and

$$\left.\begin{aligned} \partial\alpha/\partial S_1 = \partial\gamma/\partial S_1 = \partial\varepsilon/\partial S_1 = 0, \\ \partial\beta/\partial S_1 = -\,\partial\delta/\partial S_1 = 2l_xl_zXY^2, \end{aligned}\right\} \quad \text{when } S_1 = 0. \tag{10.27}$$

Hence $\partial q/\partial S_2$ and thence $\mathrm{d}y/\mathrm{d}z$, (10.17), are zero, which shows that the wave packet remains in the x–z plane, that is in the magnetic meridian.

At vertical incidence q is the same as the refractive index n. We neglect collisions so that $U = 1$ in (6.16). Then (10.17) gives

$$\frac{\mathrm{d}x}{\mathrm{d}z} = \frac{l_xl_yXY^2(n^2-1)}{2\alpha n^2 + \gamma}. \tag{10.28}$$

This is the tangent of the angle of inclination of the ray path to the vertical. By using the appropriate value of n it may be applied to either the ordinary or the extraordinary ray. Now n must satisfy $\alpha n^4 + \gamma n^2 + \varepsilon = 0$ which shows that the denominator in (10.28) is $\pm(\gamma^2-4\alpha\varepsilon)^{\frac{1}{2}}$ and

$$\frac{dx}{dz}=\pm\frac{l_x l_z Y(1-n^2)}{\{4l_z^2(1-X)^2+l_x^4Y^2\}^{\frac{1}{2}}} \tag{10.29}$$

where the minus sign applies to the ordinary ray and the plus sign to the extraordinary ray. This has the same values for upgoing and downgoing wave packets so the upward and downward paths are the same. In the northern hemisphere l_z is positive and l_x is negative so that dx/dz is positive for the ordinary ray. In the southern hemisphere l_x and l_z are both negative and dx/dz is negative for the ordinary ray. Hence the ordinary ray is always deviated towards the nearest magnetic pole and the extraordinary ray towards the equator. For frequencies greater than the electron gyro-frequency, $n_{\mathrm{O}}^2>n_{\mathrm{E}}^2$ when $X<1-Y$. Hence $(1-n_{\mathrm{O}}^2)<(1-n_{\mathrm{E}}^2)$ and the inclination of the ray path to the vertical is smaller for the ordinary ray.

The ordinary ray is reflected when $X=1$ and there $dx/dz=l_z/l_x=-\cot\Theta$ where Θ is the angle between the earth's magnetic field and the vertical. Hence at reflection the ordinary ray path is perpendicular to the earth's magnetic field, as was shown by Poeverlein's construction, § 10.9, figs. 10.4, 10.5. It should be remembered, however, that the wave normal is always vertical. It was explained in ch. 5 that the ray path and the wave normal are not in general parallel. See problem 10.7 on the polarisation of a wave at its reflection level.

Fig. 10.11. The paths of vertically incident wave packets in the northern hemisphere for a frequency greater than the electron gyro-frequency. The observer is looking towards the (magnetic) west.

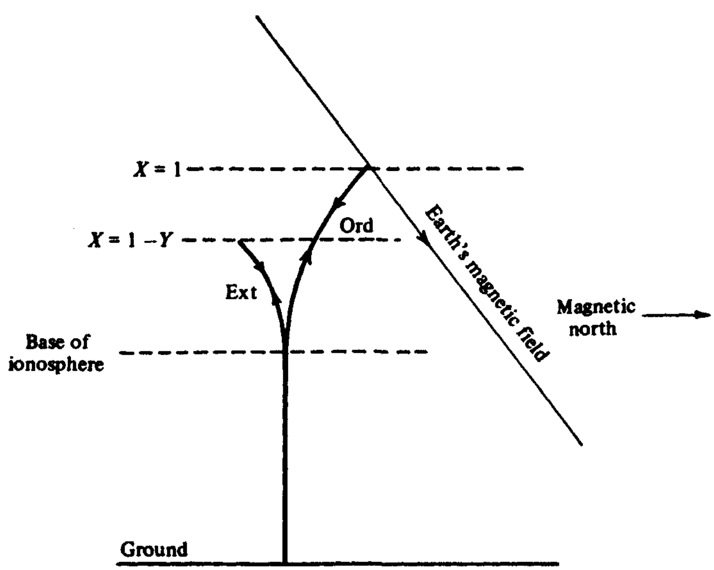


Fig. 10.11 shows the form of the two ray paths for vertical incidence. An important result of this lateral deviation is that the reflection points for the ordinary and extraordinary rays are at different latitudes. The gyro-frequency is sometimes assessed by measuring the difference of the penetration frequencies of an ionospheric layer for the two rays. If the electron density at the maximum of the layer varies appreciably with latitude the measurement will give a wrong value for the gyro-frequency (discussed by Millington, 1949, 1951).

10.13. Lateral deviation for propagation from (magnetic) east to west or west to east

In § 6.7 it was shown that when propagation is from (magnetic) east to west or west to east, the coefficients β and δ in the Booker quartic are zero and the quartic is a quadratic equation for q^2. In finding the path of a wave packet, however, it is important to notice that β and δ must not be set equal to zero until after the differential coefficients in (10.17) have been found. This is because a wave packet may include some component plane waves whose normals are not in the x–z plane.

In (6.16) put $l_x = 0$, and $U = 1$. Then

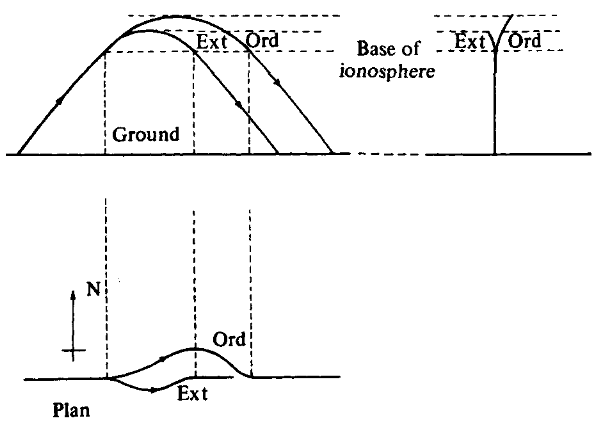
$$\frac{\partial\alpha}{\partial S_2} = 0, \frac{\partial\beta}{\partial S_2} = 2l_y l_z XY^2, \frac{\partial\gamma}{\partial S_2} = 0, \frac{\partial\delta}{\partial S_2} = -2Cl_y l_z XY^2, \frac{\partial\varepsilon}{\partial S_2} = 0, \tag{10.30}$$

when $S_2 = 0$, and (10.17) gives

$$\frac{dy}{dz} = -\left(\frac{\partial q}{\partial S_2}\right)_{S_2=0} = \frac{l_y l_z SY^2(q^2 - C^2)}{2\alpha q^2 + \gamma}. \tag{10.31}$$

This shows that the wave packet is deviated out of the x–z plane. If its path is projected on to the y–z plane then dy/dz is the tangent of the angle of inclination of

Fig. 10.12. The paths of wave packets in the northern hemisphere when the plane of incidence is (magnetic) west to east.



the projected path to the vertical. By using the appropriate value of q, (10.31) may be applied to either the ordinary or the extraordinary ray. It is the same for an upgoing or downgoing wave packet so that the upward and downward projected paths are the same. Thus a wave packet may leave the x–z plane on its upward journey but returns to it again on its downward journey. This is illustrated in fig. 10.12 which shows the projection of the ray paths on to the horizontal plane and on to two vertical planes.

Now q must satisfy the equation $\alpha q^4 + \gamma q^2 + \varepsilon = 0$ which shows that the denominator in (10.31) is $\pm(\gamma^2 - 4\alpha\varepsilon)^{\frac{1}{2}}$, and

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \pm \frac{l_y l_z Y^2(C^2 - q^2)}{\{Y^4(1 - C^2 l_z^2)^2 + 4Y^2 l_z^2(1-X)(C-X)\}^{\frac{1}{2}}}. \tag{10.32}$$

This is similar to (10.29) to which it reduces when $C = 1$ (with y for x and l_y for l_x). (See Millington, 1951.)

10.14. Lateral deviation in the general case

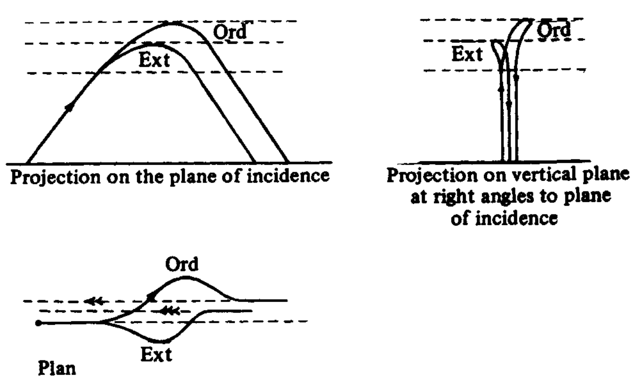
In the general case the derivatives with respect to S_2 of the coefficients (6.16) of the Booker quartic, with $U = 1$, are

$$\left.\begin{aligned} &\frac{\partial \alpha}{\partial S_2} = 0, \quad \frac{\partial \beta}{\partial S_2} = 2 l_y l_z X Y^2, \quad \frac{\partial \gamma}{\partial S_2} = 2 l_x l_y S X Y^2, \\ &\frac{\partial \delta}{\partial S_2} = -2C^2 l_y l_z X Y^2, \quad \frac{\partial \varepsilon}{\partial S_2} = 2C^2 l_x l_y S X Y^2, \end{aligned}\right\} \tag{10.33}$$

when $S_2 = 0$, where S is written for S_1. The formulae (10.17) now give

$$\frac{\mathrm{d}y}{\mathrm{d}z} = -\left(\frac{\partial q}{\partial S_2}\right)_{S_2=0} = \frac{2 l_y X Y^2(q^2 - C^2)(l_z q + l_x S)}{4\alpha q^3 + 3\beta q^2 + 2\gamma q + \delta}. \tag{10.34}$$

Fig. 10.13. The paths of wave packets in the northern hemisphere when the plane of incidence is not in the direction of one of the four (magnetic) cardinal points. The double and triple arrows show the apparent direction of the transmitter when using the ordinary and extraordinary rays respectively.



If the ray path is projected on the y–z plane then $\mathrm{d}y/\mathrm{d}z$ is the tangent of the angle of inclination of the projected path to the vertical. Now (10.34) contains odd powers of q and, at a given level, is not the same for the upward and downward paths. Hence the projections of the upward and downward paths are different. A wave packet leaves the x–z plane on its upward path and does not return to it again, or crosses it and returns to earth in a different plane. Below the ionosphere $\mathrm{d}y/\mathrm{d}z$ is zero, so that a wave packet returns to earth in a plane parallel to the x–z plane but not necessarily coincident with it. The only exceptions to this are for propagation from (magnetic) north to south or south to north, when the wave packet always remains in the x–z plane, and for propagation from (magnetic) east to west or west to east, when the wave packet moves out of the x–z plane but returns to it when it leaves the ionosphere.

The form of the ray paths in the general case is illustrated in fig. 10.13. The lateral deviation in the general case may mean that a signal reaching a receiver from a sender which is not in the direction of one of the four (magnetic) cardinal points, arrives in a vertical plane different from that containing the receiver and sender. The wave normals of this signal are in the plane of arrival, so that if a direction-finding aerial is used, the apparent bearing of the sender will differ from the true bearing. This theory has been discussed by Booker (1949) and Millington (1954), who showed that the bearing error would not be serious except in unusual conditions.

10.15. Calculation of attenuation, using the Booker quartic

Although this chapter is mainly concerned with loss-free media, so that q is purely real at all points of a ray, it is here of interest to consider the effect of a small electron collision frequency ν. The quantity q depends on ν only through $U = 1 - \mathrm{i}Z$. Let q_0 be its value when $Z = 0$. Then

$$q = q_0 - \mathrm{i}Z\frac{\partial q}{\partial U} - \frac{1}{2}Z^2\frac{\partial^2 q}{\partial U^2} + \cdots, \tag{10.35}$$

where the derivatives have their values for $Z = 0$. It can easily be shown that $\partial q/\partial U$ is then real when q is real. If Z is small, only two terms in (10.35) need be used. Then the real part of q is the same as if there were no collisions, and determines the path of the wave packet, which is the same as before. Alternatively the third term of (10.35) may be retained. This gives a slightly changed value of $\mathrm{Re}(q)$ so that the path of the wave packet is modified. There is now an imaginary part $-\mathrm{i}Z(\partial q/\partial U)$ of q, which gives attenuation of the wave. Equation (10.4) shows that the amplitude is reduced by a factor

$$\exp\left\{-k\int_0^z Z\frac{\partial q}{\partial U}\mathrm{d}z\right\}, \tag{10.36}$$

where the range of integration extends over the path of the wave packet. For a ray reflected from the ionosphere the range would extend from the ground to the level of reflection and back to the ground, and the integrand would in general be different for the upward and downward paths.

The quantity $\partial q/\partial U$ can be found from the quartic (6.15) using the form (6.16) for the coefficients. Since the quartic is satisfied for all values of U, we have $\mathrm{d}F(q)/\mathrm{d}U = 0$, whence

$$\frac{\partial q}{\partial U} = -\frac{q^4\dfrac{\partial\alpha}{\partial U} + q^2\dfrac{\partial\gamma}{\partial U} + \dfrac{\partial\varepsilon}{\partial U}}{4q^3\alpha + 3q^2\beta + 2q\gamma + \delta} \tag{10.37}$$

(β and δ do not contain U and therefore do not appear in the numerator). Booker (1949) has given curves showing how $\partial q/\partial U$ depends on X in some special cases. For another method of finding the attenuation see §§14.7, 14.10.

In § 10.5 it was shown that when every component wave in a wave packet is reversed in direction, the wave packet simply retraces its original path back to the transmitter. At each point of the path q, S_1 and S_2 are reversed in sign. Now (10.37) shows that $\partial q/\partial U$ is also simply reversed in sign, but the path of integration used in determining the attenuation is also reversed so that the total attenuation is the same for the original and the reversed paths. This result has an important bearing on the study of reciprocity (§§ 14.13, 14.14).

10.16. Phase path. Group or equivalent path

Let the vector $\mathrm{d}\boldsymbol{s}$, with components $\mathrm{d}x, \mathrm{d}y, \mathrm{d}z$, be an element of the path of a ray, and let the predominant wave in the ray have a refractive index vector $\boldsymbol{n}$ with components S_1, S_2, q (compare (10.4); the second subscripts 0 are now omitted). The change of phase in the path element $\mathrm{d}\boldsymbol{s}$ is

$$k\boldsymbol{n}\cdot\mathrm{d}\boldsymbol{s} = k(S_1\,\mathrm{d}x + S_2\,\mathrm{d}y + q\,\mathrm{d}z). \tag{10.38}$$

Suppose the ray runs from the origin to the point x, y, z. In this section of its path the change of phase is

$$k\int \boldsymbol{n}\cdot\mathrm{d}\boldsymbol{s} = kP = k\left(S_1x + S_2y + \int_0^z q\,\mathrm{d}z\right) \tag{10.39}$$

and P is called the phase path. Thus P/c is the time it would take a feature of the wave, such as a wave crest, to travel from the origin to the point x, y, z. The time of travel P'/c of a wave packet over the same path was found in § 10.3 and given by (10.11), whence

$$P' = \partial(fP)/\partial f = \partial(kP)/\partial k \tag{10.40}$$

($k = 2\pi f/c$). A wave packet travels with the group velocity and so P' is called the group path, or equivalent path.

The formulae (10.39) and (10.11), (10.40) are used in this book mainly for vertically incident waves, §§ 13.3–13.6, and for oblique and vertical incidence in an isotropic ionosphere, §§ 12.2–12.8. Some results from (10.39) for oblique incidence and an anisotropic ionosphere have been given by Booker (1949).

For a medium with losses, q is complex as shown in § 10.15, and so P, (10.39), is also complex. Then (10.36) shows that the amplitude of the signal is reduced by a factor

$$\exp\{k\,\mathrm{Im}(P)\} \tag{10.41}$$

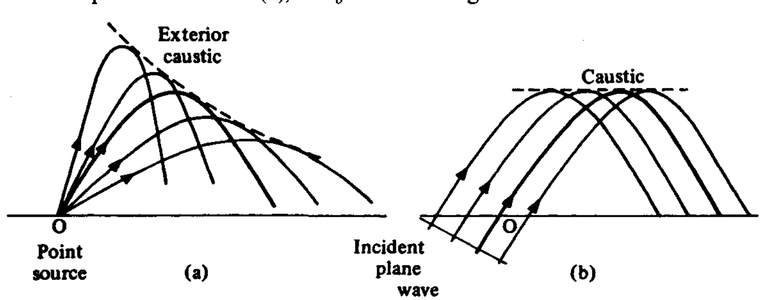
as it travels along the ray. The group path P' is also complex. The physical significance of this is discussed in § 11.16.

10.17. Ray pencils

The equations (10.3) give the path of one ray in a stratified medium. But one ray cannot exist in isolation. At any point on it there must be a wave field of the type (10.4) which extends laterally, and at nearby points not on the original ray there must be other rays. Thus any ray is just one of a family or pencil of rays. The intensity of the signal on a ray depends on the convergence or divergence of neighbouring rays, but the ray tracing equations such as (10.3) give no information about this. For example in fig. 10.14 the ray shown as a thick line is exactly the same in the two figures, but the neighbouring rays have very different configurations.

Within a family of rays we can select a set that lie in the surface of a narrow tube. Since, in a loss-free medium, the energy flux at each point is in the direction of the ray, it follows that no energy can flow across the surface of such a tube. For a continuous wave system with a fixed single frequency the energy flux does not vary with time. Then the power crossing any cross section of a ray tube must be the same for all cross sections along its whole length. This property can be used to find how the signal intensity varies along a ray path. The intensity is affected by convergence or divergence of neighbouring rays in a ray pencil. The effect of convergence is

Fig. 10.14. Typical families of rays entering the ionosphere from below. In (a) the transmitter is a point source at O. In (b) the incident wave is a plane wave in the free space below the ionosphere. In practice the true situation is usually more complicated than in (a); see § 10.21 and fig. 10.16.



important near a focus, or near the edge of a 'skip' zone, § 10.22, or near a caustic surface, § 10.18. Various aspects of the problem have been studied, mainly for loss-free media, by Försterling and Lassen (1933), Inston and Jeffs (1968), Kelso (1964) and Rawer (1948, 1952), who gives some further references.

Equations (10.5) show that the direction cosines of a ray are proportional to

$$-\partial q/\partial S_1, \quad -\partial q/\partial S_2, \quad 1. \tag{10.42}$$

Now let

$$K_{ij} = \int_0^z \frac{\partial^2 q}{\partial S_i \partial S_j} \mathrm{d}z \quad (i,j = 1,2), \tag{10.43}$$

$$\Delta = \det(K_{ij}).$$

It will be necessary to apply these formulae both to upgoing rays and to downgoing rays that have been reflected at a higher level. It has been shown in §§ 8.20, 7.19 that for these downgoing rays the integral in (10.4) must be regarded as extending from $z = 0$ up to the reflection level $z = z_0$ and thence down to the level z; see (7.153), (7.155). This same extension is implied in the integral of (10.43) when it is used for downgoing rays.

Consider a narrow ray pencil of rectangular cross section emerging from the origin into free space, and for the four rays at its corners let S_1 and S_2 be given by (S_1, S_2), $(S_1 + \delta S_1, S_2)$, $(S_1, S_2 + \delta S_2)$, $(S_1 + \delta S_1, S_2 + \delta S_2)$. Let $\mathscr{N}_0$ be the power per unit solid angle in the pencil. Then the power flowing in the pencil is

$$\mathscr{N}_0 \delta S_1 \delta S_2 / C. \tag{10.44}$$

Since the medium is loss-free, this is the same for all cross sections of the pencil.

The pencil enters the ionosphere and cuts any plane $z = \text{constant}$, in a parallelogram whose sides are, from (10.3), (10.43), the two vectors in this plane with components

$$\left.\begin{aligned} \delta x = K_{11}\delta S_1, \quad \delta y = K_{21}\delta S_1, \\ \text{and} \qquad \delta x = K_{12}\delta S_2, \quad \delta y = K_{22}\delta S_2. \end{aligned}\right\} \tag{10.45}$$

Thus the area of the parallelogram is

$$\delta S_1 \delta S_2 \Delta \tag{10.46}$$

and the power (10.44) crosses this area obliquely in the direction given by (10.42). The time average of the energy flux within the pencil (power per unit area) in the direction of the ray is

$$\begin{aligned} \Pi &= \frac{\mathscr{N}_0 \delta S_1 \delta S_2}{C} \Bigg/ \frac{\delta S_1 \delta S_2 \Delta}{\{1 + (\partial q/\partial S_1)^2 + (\partial q/\partial S_2)^2\}^{\frac{1}{2}}} \\ &= \mathscr{N}_0 \{1 + (\partial q/\partial S_1)^2 + (\partial q/\partial S_2)^2\}^{\frac{1}{2}} / C\Delta. \end{aligned} \tag{10.47}$$

This is the required result, that allows for the divergence or convergence of neighbouring rays in a pencil. If the medium has small losses, (10.47) must be multiplied by the square of (10.41).

For an isotropic plasma q is given by (10.12). We can take $S_2 = 0$ without loss of generality. Then (10.43), with (10.14) gives

$$K_{11} = -\int_0^z \frac{n^2}{q^3}\mathrm{d}z, \quad K_{12} = K_{21} = 0, \quad K_{22} = -\int_0^z \frac{\mathrm{d}z}{q} = -x/S_1 \qquad (10.48)$$

and (10.47) is

$$\Pi = \mathcal{N}_0 n/(qCK_{11}K_{22}) = \mathcal{N}_0 nS_1/(-xqCK_{11}). \qquad (10.49)$$

The expressions (10.47), (10.49) are functions of the coordinates x, y, z. They can be infinite in certain conditions, and then

$$\Delta = 0, \quad \text{or} \quad K_{11} = 0 \qquad (10.50)$$

respectively. These are the equations of a surface which is the envelope of the system of rays leaving the origin. It is not difficult to prove this directly by differential geometry. At any point on this envelope, the rays in a pencil have converged to a line or point focus. Near here the methods of ray theory fail and a more elaborate full wave treatment is needed to study the signal intensity. Some examples are given in the following section.

The source may not always be at the origin. An important case is when a plane wave is obliquely incident from free space, as would occur from an infinitely distant point source. The ray pencil at the origin is then formed from parallel rays. Let the power per unit area in the direction of the ray here be Π_1. This pencil now cuts any plane z = constant in an area that is independent of z, so the energy flux in the direction of the ray is, from (10.42)

$$\Pi = \Pi_1 C\{1 + (\partial q/\partial S_1)^2 + (\partial q/\partial S_2)^2\}^{\frac{1}{2}}. \qquad (10.51)$$

For an isotropic medium this reduces to

$$\Pi = \Pi_1 nC/q. \qquad (10.52)$$

10.18. Caustics

A surface forming the envelope of a system of rays is called a caustic surface. An example was given in § 8.22. The objective of the present section is to study how the signal from a radio transmitter varies in space when a caustic surface is present. The theory is given only for an isotropic ionosphere. The theory for an anisotropic ionosphere has not yet been fully worked out. The following version gives only the main points of the theory. For more details see Budden (1976), Maslin (1976a, b). The subject has also been studied by Kravtsov (1964a, b) Ludwig (1966), and Ziolkowski and Deschamps (1984).

Fig. 10.14(b) shows a situation where the incident wave is plane, and all rays have the same shape and the same value of S_1. They all touch a horizontal plane at the 'reflection level' where $q = 0$ and the rays are horizontal. This plane is the caustic surface. According to (10.52) the signal intensity would here be infinite, but it has been shown by a 'full wave' study in § 8.20 that the height dependence of the signal is actually given by an Airy integral function, as in (8.72). Above it, although there are no real rays, the signal does not fall abruptly to zero but decays with increasing height in accordance with (7.65). The field here is an inhomogeneous wave, § 2.15, and can be thought of as containing complex rays; see § 14.11. The region where the rays are real is called the 'illuminated' side of the caustic, and the region where there are no real rays is called the 'dark' side.

Fig. 10.14(a) is an example where the rays come from a point transmitter at the origin. They have various different values of S_1 and therefore become horizontal at different levels, in each case where $q = 0$. Their envelope is shown as a broken line in the figure but it is actually the cross section by the plane $y = 0$ of a surface, the caustic surface. In this example it is called an exterior caustic because it is an outer boundary of the region of space where the rays are real. Another type, called an interior caustic, is possible and discussed in § 10.21, fig. 10.16.

For a point source at the origin the emitted field is given by the angular spectrum of plane waves (10.1). Here the exponential is an obliquely upgoing wave of unit amplitude. A downgoing wave reflected from higher levels may also be present but is not included in (10.1). Suppose, now, that the transmitter is a vertical Hertzian magnetic dipole of unit amplitude. Then the electric field is everywhere horizontal. In the plane $y = 0$ studied here, the only non-zero electric field component is E_y. If this is chosen for F in (10.1), it can be shown (Clemmow, 1966, § 2.25) that

$$A(S_1, S_2) = S_1/C \tag{10.53}$$

where a constant multiplier $ick^3/8\pi^2$ has been omitted. Similarly, if the transmitter is a vertical Hertzian electric dipole, the only non-zero magnetic field component in the plane $y = 0$ is $\mathscr{H}_y$. If $-\mathscr{H}_y$ is chosen for F in (10.1), then (10.53) is again true. Other types of transmitter and other field components can be dealt with by using the appropriate A in (10.53). In the following theory F can be interpreted either as E_y or as $-\mathscr{H}_y/n$. The need for the factor $1/n$ here is shown by (7.74).

It is now assumed that in the ionosphere the electron concentration $N(z)$ increases monotonically with height, so that each plane wave in (10.1) is reflected. The case where some waves penetrate an ionospheric layer is considered later, § 10.22. The exponential in (10.1) must be replaced by the uniform approximation (8.72) for one component plane wave. This gives, with $y = 0$,

$$F = \iint 2S_1\zeta^{\frac{1}{4}}(\pi/Cq)^{\frac{1}{2}}\mathrm{Ai}(\zeta)\exp\left\{-ik\left(S_1 x + \int_0^{z_0} q\,\mathrm{d}z\right)\right\}\mathrm{d}S_1\,\mathrm{d}S_2 \tag{10.54}$$

where ζ is given by (8.68). If the asymptotic form for $\mathrm{Ai}(\zeta)$ is used, this gives

$$F \approx \iint S_1 (Cq)^{-\frac{1}{2}} \exp(-\mathrm{i}kS_1 x)\left\{\exp\left(-\mathrm{i}k\int_0^z q\,\mathrm{d}z\right)\right.$$
$$\left. + \mathrm{i}\exp\left(-\mathrm{i}k\oint_0^z q\,\mathrm{d}z\right)\right\}\mathrm{d}S_1\,\mathrm{d}S_2. \tag{10.55}$$

The first term is the upgoing wave, and the integral in the exponent goes from 0 up to z. The second term is the reflected wave. The symbol $\oint$ implies that the integral extends from $z = 0$ to $z = z_0$ where $q = 0$ and the ray is horizontal, and thence down to the level z with the sign of q reversed, as in (7.153). A caustic surface is nearly always an envelope of the downgoing rays, as in the example of fig. 10.14(a). Near it, therefore, the second term of (10.55) usually has to be studied. But here we shall study both terms. Note that the factor i from (8.73) is retained in (10.55).

The S_1 and S_2 integrals in (10.54), (10.55) are now to be evaluated by the method of steepest descents, ch. 9. We study first (10.55). Either of its two terms may be written

$$F = \iint S_1 (Cq)^{-\frac{1}{2}} \exp(-\mathrm{i}kp)\mathrm{d}S_1\,\mathrm{d}S_2, \quad p = S_1 x + \int_0^z q\mathrm{d}z \tag{10.56}$$

and the integral sign in p may be either $\int$ or $\oint$. The double integral for F requires the method of double steepest descents, §9.10, but the process is simple in this case because the saddle point for the S_2 integration is at $S_2 = 0$ and independent of S_1. The second derivative $\partial^2 p/\partial S_1 \partial S_2$ is zero. Thus the S_2 integration is done first. It uses

$$\frac{\partial^2 p}{\partial S_2^2} = K_{22} = -x/S_1 \tag{10.57}$$

from (10.43), (10.48). This integration is perhaps more conveniently thought of as using the method of stationary phase; compare §10.2, especially (10.3) second equation. The result is, from (9.31)

$$F = (2\pi)^{\frac{1}{2}} \exp\left(\tfrac{1}{4}\mathrm{i}\pi\right)\int S_1^{3/2}(kxCq)^{-\frac{1}{2}} \exp(-\mathrm{i}kp)\,\mathrm{d}S_1 \tag{10.58}$$

where C, q, p now take their values for $S_2 = 0$.

Finally the S_1 integration must be done. When the saddle points in the S_1 plane are well separated it is sufficiently accurate to use the first order steepest descents result (9.31) at each. This is equivalent to using ray theory. Where two saddle points coincide there is a caustic and for points near it an extension of the method of steepest descents is needed, as in § 10.19. Where three saddle points coincide there is a cusp; see § 10.21. These properties were used by Maslin (1976a, b) to trace the caustic curves in the plane $y = 0$ and to find the cusps. He showed that, for a monotonic electron height distribution $N(z)$ and for the first term in (10.55), that is for upgoing waves, there are no caustics or cusps. These occur only for the second term, that is for downgoing waves. They can occur, however, for upgoing waves that have penetrated a lower ionospheric layer. The number of contributing saddle points in (10.58) depends on the coordinates x, z. Maslin (1976a, b) showed that it can be as many as five in practical cases.

We now find the contribution to (10.58) from one isolated saddle point. It must occur where $\partial p/\partial S_1 = 0$ which gives $x = S_1 \int_0^z q^{-1}\,\mathrm{d}z$. This is the same as (10.14). It shows that the saddle point value of S_1 is the value for the ray through the point (x, z). Let this value be S_a and let $q = q_a$, $C = C_a$ for $S_1 = S_a$. The second derivative of p when $S_1 = S_a$ is, from (10.43), (10.49)

$$\partial^2 p/\partial S_1^2 = K_{11} = -\int_0^z (n^2/q_a^3)\,\mathrm{d}z. \tag{10.59}$$

Then the contribution to (10.58) from the saddle point is, from (9.31):

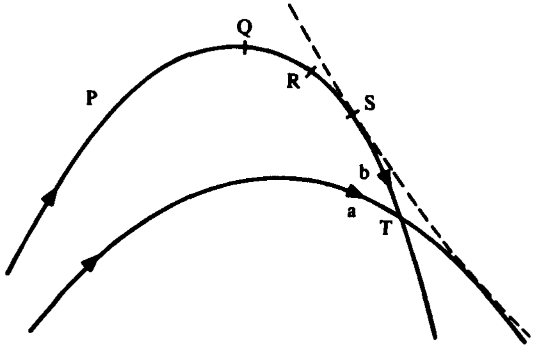
$$F = 2\pi \mathrm{i} k^{-1} S_a^{\frac{3}{2}} (-C_a q_a K_{11} x)^{-\frac{1}{2}} \exp\left\{-\mathrm{i}k\left(S_a x + \int_0^z q_a\,\mathrm{d}z\right)\right\}. \tag{10.60}$$

This is the field F on a ray that is not near a caustic.

A typical situation is sketched in fig. 10.15. The curve PQRST is a ray that touches the caustic at S. At P it is obliquely upgoing so that (10.60) uses the first term of (10.55). Here q_a is real and positive and (10.59) shows that K_{11} is real and negative. Thus the factor $(-q_a K_{11})^{-\frac{1}{2}}$ in (10.60) is real and positive. For the second derivative (9.27) of the exponent, $\alpha = \frac{1}{2}\pi$ so that the last factor of (9.29) is $\exp(\frac{1}{4}\mathrm{i}\pi)$. This has been combined with the same factor in (10.58) to give the factor i in (10.60).

The approximation for $\mathrm{Ai}(\zeta)$ used in (10.55) fails near $q = 0$ so that it might be expected that (10.60) fails at Q where $q_a = 0$. Here (10.59) shows that K_{11} is infinite. Although q_a and K_{11} both have branch points where $q_a = 0$, it is easy to show that the product $q_a K_{11}$ is analytic and non-zero there. When Q is passed and we move to R, both q_a and K_{11} have changed sign so q_a is negative and K_{11} is positive. Now in

Fig. 10.15. The continuous curves are two of a family of rays that touch a caustic surface, shown as a broken line.



(9.27) α is $3\pi/2$ and the last factor of (9.31) is $\exp(-\frac{1}{4}i\pi)$. But now the ray is obliquely downgoing and the factor i from the second term of (10.55) must be included. This finally gives a factor $\exp(+\frac{1}{4}i\pi)$ as before, and (10.60) is still correct for points such as R. It is continuous for points near Q and it turns out that it is still correct there, in spite of the failure of (10.55). This can be proved by the method outlined in § 10.19.

At S in fig. 10.15, two saddle points S_a coincide. This requires that $\partial^2 p/\partial S_1^2 = 0$ so that $K_{11} = 0$ from (10.59); compare (10.50). But q_a is bounded so (10.60) is infinite. Here the simple form of steepest descents fails, so that (10.60) is not valid even though (10.55) is a good approximation.

At T on the same ray K_{11} is again negative. The result (10.60) is valid provided that the extra factor i from the second term of (10.55) is included. It is not now cancelled because of the last term of (9.29). It represents the $\frac{1}{2}\pi$ phase advance that occurs in a ray that has previously touched a caustic surface.

10.19. The field where the rays are horizontal

When the approximation used for (10.55) does not apply, that is near $q = 0$, the more accurate form (10.54) must be used. The algebra for this case, given in full by Budden (1976), is lengthy and only the main results are given here.

For the function Ai in (10.54) we use its integral representation (8.16). This gives

$$F = -i\int\int\int S_1\zeta^{\frac{1}{4}}(\pi Cq)^{-\frac{1}{2}}\exp\left\{-ik\left(S_1x + \int_0^{z_0} q\,dz\right) + \zeta t - \frac{1}{3}t^3\right\}dS_1\,dS_2\,dt. \tag{10.61}$$

This is a triple integral but, as for (10.56), the S_2 integration is simple and can be done first by the method of stationary phase; compare (10.58). The result is

$$F = \exp(-\tfrac{1}{4}i\pi)(2/kx)^{\frac{1}{2}}\int\int S_1^{3/2}\zeta^{\frac{1}{4}}(Cq)^{-\frac{1}{2}}\exp(-ikP + \zeta t - \tfrac{1}{3}t^3)dS_1\,dt \tag{10.62}$$

where

$$P = S_1x + \int_0^{z_0} q\,dz. \tag{10.63}$$

The integral (10.62) is done by the method of double steepest descents, §9.10. The double saddle points are where $\partial/\partial S_1$ and $\partial/\partial t$ of the exponent are zero. It can be shown that these lead to the first equation (10.14) for finding S_1. This therefore has the value S_a for the ray that goes through the point (x, z).

If the double saddle points are well separated in S_1–t space, the first order double steepest descents formula (9.64) can be used for each. It can thence be shown that the contribution from one double saddle point is (10.60). This result now applies even for a point such as Q in fig. 10.15 where the ray is horizontal. It supplies the proof mentioned near the end of § 10.18.

The W.K.B. solutions of ch. 7 were for the fields in a stratified ionosphere when a plane wave was incident from below it. These solutions failed at a level where the rays were horizontal and touched a horizontal caustic surface. In (10.60) we now have a very important new result. It is a generalisation of the W.K.B. solution and applies when the waves originate from a point source not at infinity. Like the conventional W.K.B. solution it is approximate because its derivation used the first order steepest descents formulae. It fails where the rays touch a caustic surface. But it remains valid where the rays are horizontal provided this occurs at points not near a caustic.

10.20. The field near a caustic surface

Near a caustic surface the integrand of (10.62) has two double saddle points close together so the first order double steepest descents formula cannot be used. For a single integral such as (9.1) with two saddle points close together, the method of Chester, Friedman and Ursell (1957), §9.9, can be used. For the double integral (10.62) an extension of this has been suggested (Budden, 1976). The full theory has been discussed by Ursell (1980). The results are as follows.

Consider a point such as T in fig. 10.15 where two rays intersect. Let subscripts a, b be used to indicate values at T from these two rays, a for the ray that has not yet touched the caustic and b for the ray that has touched it. Thus P_{a}, P_{b} are the two phase paths, from (10.39) with $y = 0$, and S_{a}, S_{b} are the two values of S_1, etc. Now define

$$\xi = \{\tfrac{3}{4}\mathrm{i}k(P_{\mathrm{a}} - P_{\mathrm{b}})\}^{\frac{2}{3}} \tag{10.64}$$

where the fractional power is chosen so that ξ is real and negative when $P_{\mathrm{b}} - P_{\mathrm{a}}$ is real and positive. This should be compared with (8.68) for ζ. If the source is at infinity so that the incident rays a and b are parallel and $S_{\mathrm{a}} = S_{\mathrm{b}}$, then $\xi = \zeta$. Let

$$g_{\mathrm{a}} = \mathrm{i}k^{-1}S_{\mathrm{a}}^{3/2}(-C_{\mathrm{a}}q_{\mathrm{a}}K_{11\mathrm{a}}x)^{-\frac{1}{2}}, \quad g_{\mathrm{b}} = \mathrm{i}k^{-1}S_{\mathrm{b}}^{3/2}(C_{\mathrm{b}}q_{\mathrm{b}}K_{11\mathrm{b}}x)^{-\frac{1}{2}}. \tag{10.65}$$

$$V = \xi^{\frac{1}{4}}(g_{\mathrm{b}} + g_{\mathrm{a}}), \quad W = \xi^{-\frac{1}{4}}(g_{\mathrm{b}} - g_{\mathrm{a}}). \tag{10.66}$$

Note that the factor g_{a} appears in (10.60). Then it can be shown that

$$F = 2\pi^{\frac{1}{2}}\{V\,\mathrm{Ai}(\xi) + W\,\mathrm{Ai}'(\xi)\}\exp\{-\tfrac{1}{2}\mathrm{i}k(P_{\mathrm{a}} + P_{\mathrm{b}})\}. \tag{10.67}$$

If the point T is beyond the caustic in fig. 10.15, there are no real rays through it. But the quantities S_{a}, S_{b}, P_{a}, P_{b}, etc. can still be defined. They are now complex and we say that complex rays can reach points beyond the caustic; see §14.11. It can be shown that ξ is a continuous real function for points x, z on both sides of the caustic. It is zero on the caustic and positive for points beyond it. Thus (10.67) is uniformly valid in this region. It is the analogue of the 'uniform approximation' solution (8.72) that applied for an incident plane wave.

Finally, for a point on the illuminated side of the caustic where rays are real and ξ is negative, suppose that $|\xi| \gtrsim 1$ and use the asymptotic forms (8.53), (8.55) for Ai, Ai′

in (10.67), and use (10.64) for ξ. Then

$$F \approx 2\pi\{g_a \exp(-ikP_a) + ig_b \exp(-ikP_b)\}. \tag{10.68}$$

The first term is the same as (10.60) for ray a and the second is (10.60) for ray b. The factor i shows the phase advance of $\frac{1}{2}\pi$ already explained at the end of § 10.18.

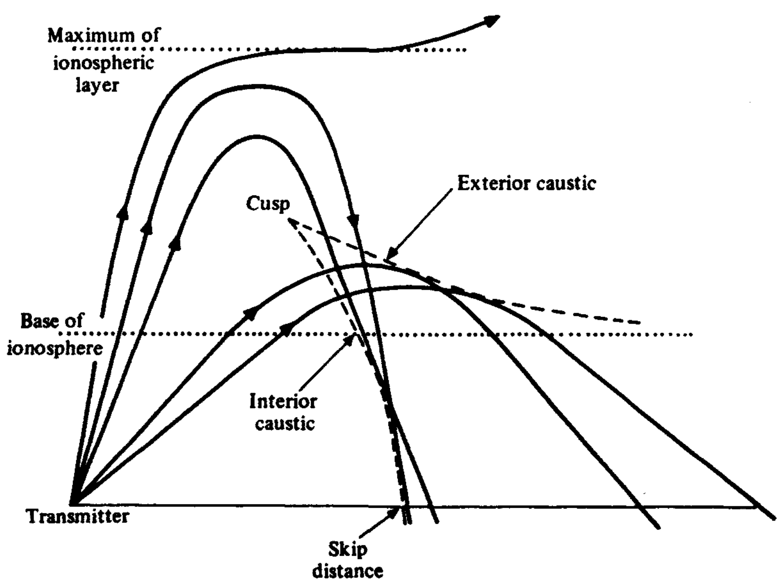
For a point on the dark side of the caustic the asymptotic forms (8.52), (8.54) would be used, and there would be only one term in (10.68). It would represent the inhomogeneous wave mentioned in § 10.18, and its amplitude would decrease rapidly with increasing distance form the caustic.

For a discussion of the above theory and for results of some calculations closely related to it, see Warren, de Witt and Warber (1982).

10.21. Cusps. Catastrophes

When a family of rays enters the ionosphere from below, the configuration is often more complicated than implied in fig. 10.14(a). A typical situation is sketched in fig. 10.16. The caustic surface is in two parts. One part is the same as the exterior caustic of fig. 10.14(a). But there is another branch which is touched by the rays on the side remote from the transmitter, and it is therefore called an interior caustic. The two branches meet in a point in fig. 10.16, but this is a cross section. The caustics are actually surfaces and they meet in a line, called a cusp line. It can be shown that near this line three saddle points of the integral (10.58) are close together. Both the simple

Fig. 10.16. Example of rays from point transmitter entering the ionosphere and shown as continuous curves. Their envelope, shown as a broken curve, is a caustic with interior and exterior branches meeting in a cusp.



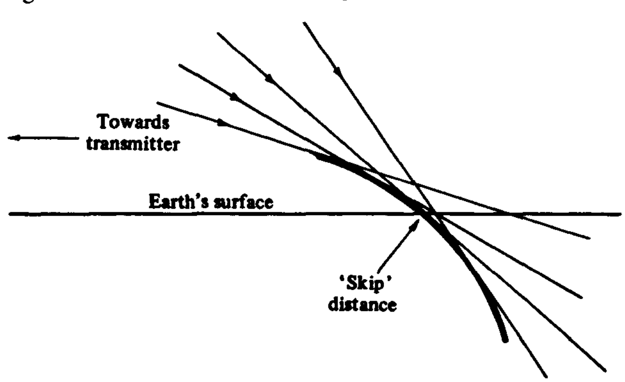
ray theory and the theory of §§ 10.18–10.20 are inadequate and a still more elaborate theory is needed. This has been given for the radio wave problem by Budden (1976). It is found that the Airy integral used near a caustic must now be replaced by a more complicated function given by Pearcey (1946), and sometimes called the 'cusp function'; (see also Maslin, 1976b). Caustics and cusps have played some part in the study of guided waves; see Hartree, Michel and Nicolson (1946), Brekhovskikh (1960).

Caustics and cusps are examples of a class of focusing phenomena that occur in physics, known as catastrophes; see, for example, Poston and Stewart (1978). They are important when ray theory is used to study the propagation of radio waves through a medium containing irregularities of refractive index. The caustic is the only catastrophe that is likely to be often encountered by practical radio engineers, and it is the only one described in any detail in this book.

10.22. The skip distance

Fig. 10.16 is typical for rays from a point transmitter entering an ionospheric layer in which the plasma frequency at the maximum is less than the frequency of the wave. Thus rays near vertical incidence penetrate the layer. If a ray is to return to the ground it must be obliquely incident. It follows that there is a minimum distance D_s from the transmitter within which no ray travelling via the ionosphere can reach the ground. Fig. 10.16 shows that this distance D_s is where the interior caustic meets the ground. Near it we expect the signal amplitude to be given by an Airy integral function, as in (10.67). This distance is called the 'skip' distance. Just beyond it two rays reach the ground and interact to give an interference pattern, expressed by the oscillations of the Airy integral function. These rays are in free space and are straight and the caustic is curved. It was shown in §8.22 that the signal amplitude along a line

Fig. 10.17. Configuration of downcoming rays near the skip distance. The thick curve is the interior caustic. The rays would actually be reflected by the earth, but in this diagram the effect of the earth is ignored.



at right angles to the caustic is proportional to

$$\mathrm{Ai}(\zeta) \text{ with } \zeta = 2u(\pi^2/R\lambda^2)^{\frac{1}{3}} \tag{10.69}$$

where u is distance measured perpendicular to the caustic, λ is the wavelength and R is the radius of curvature of the caustic.

The configuration of the rays reaching the ground near the skip distance D_s is sketched in fig. 10.17. Let the particular ray that reaches the gound at range D_s make an angle θ_0 with the vertical. Rays at neighbouring angles θ all reach the ground at greater distances D. It can easily be shown that $R = (\mathrm{d}^2D/\mathrm{d}\theta^2)\cos\theta_0$. If x is distance measured along the ground from the skip distance, towards the transmitter, then $u = x\cos\theta_0$ and the expression in (10.69) for ζ is

$$\zeta = 2x\left(\frac{\pi^2\cos^2\theta_0}{\lambda^2\mathrm{d}^2D/\mathrm{d}\theta^2}\right)^{\frac{1}{3}}. \tag{10.70}$$

Typical values of θ_0 and $\mathrm{d}^2D/\mathrm{d}\theta^2$ can be found from formulae given later, § 12.7, and used to assess the scale of the function $\mathrm{Ai}(\zeta)$. It is found that it is of order $\mathrm{Ai}(0.5x)$ to $\mathrm{Ai}(1.5x)$ where x is in km. Thus the scale of the interference pattern just beyond the skip distance is of order one or two kilometres.

If the receiver is at a fixed distance D from the transmitter, there is some frequency f_M that makes $D_s = D$, and it is called the 'maximum usable frequency'; see §§ 12.8–12.10. For other frequencies, x in (10.70) is $D_s - D$ and approximately

$$x = (f - f_M)\mathrm{d}D_s/\mathrm{d}f. \tag{10.71}$$

If the transmitted frequency f is changed slowly, the received signal amplitude varies as $\mathrm{Ai}(\zeta)$ where

$$\zeta = 2(f - f_M)\frac{\mathrm{d}D_s}{\mathrm{d}f}\left(\frac{\pi^2\cos^2\theta_0}{\lambda^2\mathrm{d}^2D/\mathrm{d}\theta^2}\right)^{\frac{1}{3}}. \tag{10.72}$$

Some properties of this formula have been discussed by Budden (1961a, § 11.12). The value of $\mathrm{d}D_s/\mathrm{d}f$ can be estimated from formulae discussed later, § 12.8.

In practical cases $\mathrm{d}D_s/\mathrm{d}f$ is usually positive. Then if $f < f_M$ two rays reach the receiver, and ζ is negative. If f increases, the signal amplitude oscillates because of the maxima and minima of $\mathrm{Ai}(\zeta)$, which is an interference pattern formed by the two rays. Eventually f may exceed f_M and ζ becomes positive. Then the signal decays and disappears. Similar effects are observed with a fixed frequency because the ionosphere changes with time and f_M changes. The oscillatory effect in the amplitude of a signal that is about to fade right out is well known to radio operators who work near the maximum usable frequency.

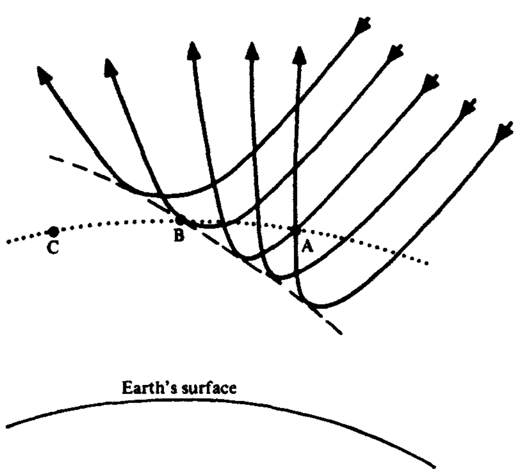
10.23. Edge focusing

A radio signal coming in from outside the earth is refracted by the ionosphere. For very high frequencies the refractive index n of the ionosphere is very close to unity and the refraction is negligible, but for smaller frequencies $|n-1|$ can be large enough to be important. The subject is of some interest in radio astronomy. We here consider only the case where the source is effectively at an infinite distance so that the rays reaching the earth from it are all parallel before they enter the ionosphere. For a study of the rays from a source at a finite distance from the earth see, for example, Rawer (1962).

In the top side of the ionosphere, that is above the maximum of the F2-layer, it can be assumed that the electron concentration $N(r)$ is a monotonically decreasing function of the radius r from the centre of the earth. The refraction of the rays is studied by ray tracing in the spherically stratified ionosphere. The frequencies of interest are so great that the earth's magnetic field and electron collisions may be neglected. The equations for the ray path were given in § 10.4. Spherical polar coordinates r, θ, ϕ are used. The receiver is assumed to be in the top side of the ionosphere on the line $\theta = 0$ at radius $r = a$. The plane $\phi = 0$ is chosen so that the direction of the source is in it, at $\theta = \theta_s$.

A ray coming in to the ionosphere is deviated away from the radial direction, and it may become horizontal and then travel outwards again. A possible situation is sketched in fig. 10.18. Here the rays all touch an external caustic surface shown as a broken line. An internal caustic could also occur but is not considered here; see fig. 10.16. For a receiver at a position such as A in fig. 10.18 two different real rays can reach it. If it moves to B, these two rays have moved to coincidence and B is on the caustic. For a receiver at C no real rays can reach it because it is in the shadow zone. Near the caustic on the illuminated side there is a concentration of rays giving a focusing effect, and this has been called 'edge focusing'.

Fig. 10.18. Rays from an infinitely distant source enter the upper ionosphere and touch a caustic surface, shown as a broken line.



A receiver in the ionosphere must be on a space vehicle that moves, so that the source direction θ_s observed at the receiver changes with time. It is required to find how the signal amplitude depends on θ_s for a fixed value of the receiver height a. The full theory for a collisionless isotropic ionosphere was given by Budden (1961c). Only an outline of the method and the result are given here.

For an inward coming ray in the ionosphere let ψ be the angle between its direction and the inward radius, and let ψ_A be the value of ψ at the receiver. Let n_0 be the refractive index at the receiver. Then Bouger's law (10.18) gives

$$rn \sin\psi = an_0 \sin\psi_a = K. \tag{10.73}$$

The differential equation of a ray path is (10.19). For a ray going through the receiver this may be integrated with respect to r to give

$$\theta_s = an_0 \sin\psi_a \,⨍_a^\infty \frac{dr}{r^2 q} \tag{10.74}$$

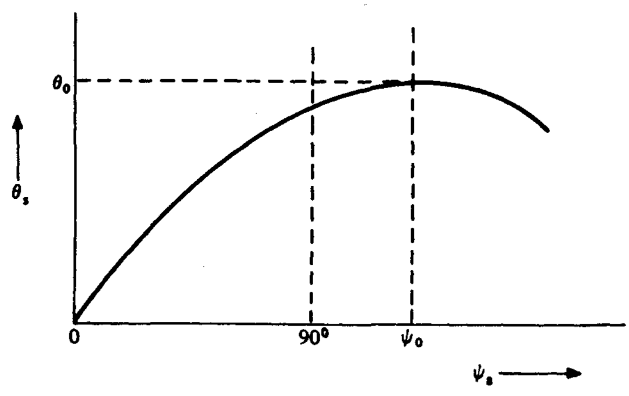
where

$$q = n\cos\psi = (n^2 - n_0^2 \sin^2\psi_a a^2/r^2)^{\frac{1}{2}}. \tag{10.75}$$

If $\cos\psi_a$ is positive, the ray at the receiver is incoming and the integral in (10.74) extends along the real r axis from a to ∞. But if $\cos\psi_a$ is negative, the ray at the receiver is outgoing, because it was horizontal where $r = r_1 < a$. The first part of the path of integration in (10.74) must therefore run from a down to r_1 and on this part $\cos\psi$ and q are negative. The second part runs from r_1 to ∞ and on it $\cos\psi$ and q are positive. The symbol ⨍ is used to indicate that the correct choice of path must be made. Some properties of the integral (10.74) were discussed by Daniell (1966) who showed that it is convenient to treat it as a contour integral in the complex r plane.

Since n^2 is a known function of r, (10.74) can be evaluated to give an equation relating θ_s and ψ_a. A specific example was given by Budden (1961c). The form of the

Fig. 10.19. Relation between direction θ_s of source, and inclination ψ_a of the ray at the receiver to the inward radius.



result is sketched in fig. 10.19. It shows that θ_s has a maximum value θ_0, which is the source direction that makes the caustic go through the receiver. It occurs where $\psi_a = \psi_0$. By differentiating (10.74) with respect to ψ_a it can be shown that $\psi_0 > \frac{1}{2}\pi$. This means that only the outgoing parts of the rays can touch the caustic. The corresponding result for a flat earth was mentioned in § 10.18 (Maslin, 1976a, b). Where θ_s is a maximum let

$$\mathrm{d}^2\theta_s/\mathrm{d}\psi_a^2 = -2\gamma. \tag{10.76}$$

Now let the source emit linearly polarised waves with the electric field E perpendicular to the plane $\phi = 0$. To find the field E in the immediate neighbourhood of the receiver Budden (1961c) assumed that, in a sufficiently small region, the earth's curvature could be neglected and that n^2 varied linearly with radius thus

$$n^2 \approx n_0^2 + \alpha(r - a). \tag{10.77}$$

The differential equation satisfied by E in this region was then formulated and solved. This showed finally that when θ_s is near to θ_0 the signal amplitude is proportional to

$$\mathrm{Ai}\{-(k/\alpha)^{\frac{2}{3}} n_0^2 (\sin\psi_0)^{\frac{1}{3}}(\theta_0 - \theta_s)/\gamma\}. \tag{10.78}$$

It was at one time suggested that edge focusing might be used for detecting a radio star source from the large signal that occurs when θ_s is near θ_0. But the maximum value of $|\mathrm{Ai}|$ is not very large; see fig. 8.5. The term 'focusing' is perhaps misleading. The method has not proved to be practicable.

PROBLEMS 10

10.1. Show that for a radio wave travelling vertically upwards in a horizontally stratified loss-free ionosphere, the time averaged Poynting vector makes with the vertical an angle whose tangent is

$$\frac{\mathrm{i}Y\rho l_x(1 - n^2)}{(1 - X)(\rho^2 - 1)}$$

(use (4.58) for a loss-free plasma). By using ρ from (4.20) show that this is the same as (10.29).

10.2. For an extraordinary ray vertically incident on a horizontally stratified loss-free ionosphere, show that at the reflection level the angle between the ray and the vertical is arctan $\{\sin 2\Theta/(3 + \cos 2\Theta)\}$ where Θ is the angle between the earth's magnetic field and the vertical.
(See Millington, 1949, 1951).

10.3. In a spherically stratified isotropic ionosphere the refractive index $n(r)$ is a function only of radius r from the earth's centre. A ray at radius r makes an angle ψ with the radius. Prove Bouger's law that $rn \sin\psi$ is constant on the ray.
(Hint: prove it first for the straight rays on the two sides of a spherical boundary

between two homogeneous media. Then extend it to a succession of boundaries and finally to the limit of a continuously varying medium.

Another method is to use the general ray tracing equations (14.33), (14.34).

A third method is to use Fermat's principle.)

10.4. A continuous radio wave from a point source on the ground is vertically incident on a horizontally stratified isotropic loss-free ionosphere. Show that at any height, where the total equivalent path travelled from the ground is P', the signal intensity is proportional to $(P')^{-2}$ for both upgoing and downgoing (reflected) waves.
(Use (10.48), (10.49)).

10.5. An ordinary ray is incident obliquely on a collisionless ionosphere, in the magnetic meridian plane at an angle of incidence arcsin $\{l_x(Y+1)^{\frac{1}{2}}\}$ (see (6.32) and § 10.9). When it reaches the level $X = 1$, it is at a window point. It continues obliquely upwards as an extraordinary ray, but there is also a reflected ordinary ray of infinitesimally small amplitude travelling downwards. (See fig. 10.8(a).) Show that these two rays are inclined to the earth's magnetic field at angles ψ given by

$$\cot\psi = 2\cot\Theta[1 + Y \pm \{(1+Y)^2 + \tfrac{1}{2}(1+Y)\tan^2\Theta\}^{\frac{1}{2}}]$$

where Θ is the angle between the earth's magnetic field and the vertical ($l_x = \sin\Theta$). What happens to these rays (a) at a magnetic pole, (b) at the magnetic equator?

10.6. A ray goes obliquely from a transmitter on the ground at (0, 0, 0) into a horizontally stratified ionosphere. It is reflected and returns to a receiver on the ground at (x, y, 0). When the ionosphere is loss-free the ray, where it leaves the transmitter, has x and y direction cosines S_1, S_2 respectively. The function $q(z)$ on the ray has a branch point at the reflection level $z = z_0$. Show that the phase path for the whole ray is

$$P = S_1 x + S_2 y + \int_{\mathrm{C}} q\,\mathrm{d}z.$$

where the contour C is as used in (8.78). When there are no collisions, S_1, S_2 and P are real and there is no attenuation. A collision frequency is now introduced so that Z is small at all heights. Then the ray path is perturbed. Prove that the attenuation, in nepers, from (10.41) is now

$$-k\,\mathrm{Im}(P) \approx k\int_{\mathrm{C}} Z\frac{\partial q}{\partial U}\mathrm{d}z.$$

This is the same as if S_1, S_2 are unchanged, and only $q(z)$ is changed. In fact the perturbed ray is complex and S_1, S_2 have complex increments $\delta S_1, \delta S_2$. The proof must take account of this.

10.7. For a radio wave travelling vertically upwards into a horizontally stratified loss-free ionosphere, show that the ordinary wave, at its reflection level, is linearly polarised with its electric vector parallel to the earth's magnetic field $\boldsymbol{B}$, and its magnetic vector horizontal and perpendicular to $\boldsymbol{B}$. Show that for the extraordinary wave, at its reflection level, the electric vector describes a circle in a plane at right angles to $\boldsymbol{B}$.

10.8. A radio ray travels in an isotropic medium whose refractive index n is a function only of the radius r measured from a fixed centre. By using Fermat's principle, or otherwise, show that the path of the ray satisfies

$$\frac{\mathrm{d}r}{\mathrm{d}\theta} = \frac{r}{K}(r^2n^2 - K^2)^{\frac{1}{2}}$$

where K is a constant and r, θ are spherical polar coordinates with respect to an axis in the plane of the ray.

For radio rays of a certain frequency, the refractive index of the earth's outer atmosphere is given by $n^2 = 1 - a^2/r^2$, where a is a constant. A ray approaches the earth along a line which, if produced, would pass at a distance b from the earth's centre. Show that the ray is deviated through an angle

$$\pi\{1 - b(a^2 + b^2)^{-\frac{1}{2}}\}.$$