14

General ray tracing

14.1. Introduction

Ch. 10 studied the problem of ray tracing in a plane stratified medium. The objective of the present chapter is to extend this theory to deal with media that are not plane stratified. For example the composition of the ionosphere nearly always depends on the horizontal coordinates, and there are cases especially near twilight when it cannot be assumed to be horizontally stratified. The structure of the magnetosphere is controlled largely by the earth's magnetic field and the solar wind, see § 1.9, and it is not plane stratified.

The number of published papers on ray tracing methods is very large and the subject has acquired some mathematical interest that often goes beyond the needs of practical radio engineers. In this book no attempt has been made to give all the references, but a few that have come to the author's attention are given at the relevant points in this chapter. The objective here is to present the basic physical ideas of ray tracing and to describe the methods that have mainly been used for radio waves in the ionosphere and magnetosphere.

In § 10.2 the ray path was expressed by integrals (10.3), and for plane stratified media these can often be evaluated in closed form. For the more general plasma, however, this is rarely possible. The ray path was also expressed by its differential equations (10.5) for $\mathrm{d}x/\mathrm{d}z$, $\mathrm{d}y/\mathrm{d}z$. The independent variable was the height z and was measured normal to the strata. In the more general case there is nothing to single out one coordinate in this way. In this chapter the ray will be traced by deriving expressions for $\mathrm{d}x/\mathrm{d}\tau$, $\mathrm{d}y/\mathrm{d}\tau$, $\mathrm{d}z/\mathrm{d}\tau$, or analogous expressions with other coordinates instead of x, y, z. Here τ is some variable that increases as we proceed along a ray. There is some latitude in its choice; see § 14.3 below. It is found that these three equations must be supplemented by three more expressions that give $\mathrm{d}/\mathrm{d}\tau$ of the components of the refractive index vector. The six differential equations for the ray path are then integrated with respect to τ. This is usually done with a computer,

 https://doi.org/10.1017/CBO9780511564321.015

using a step-by-step process. The equations will first be derived for Cartesian coordinates x, y, z. They will then be transformed to use other coordinate systems; see § 14.5.

One purpose of ray tracing is to find a ray that goes from a given transmitter to a given receiver. Suppose that the transmitter is at the origin and the receiver is on the ground at the point $(x_0, 0, 0)$. The direction of a ray where it leaves the transmitter may be specified by the polar angles θ, ϕ of its wave normal, and these are not at first known. It is therefore necessary to trace several rays and find where they return to the ground. Simplifications are often possible. For example if it is known that there is no lateral deviation of the rays, then $\phi = 0$. If then a ray that starts with $\theta = \theta_i$ returns where $x = x_i$, for $i = 1, 2$, an estimate of the correct value θ_0 of θ can be made by linear interpolation, sometimes called the 'rule of false position' (Hartree, 1958):

$$\theta_0 \approx \frac{\theta_1 x_2 - \theta_2 x_1 + x_0(\theta_2 - \theta_1)}{x_2 - x_1}. \tag{14.1}$$

To use this formula two rays must first be traced. It may often be necessary to trace two or three more before one is found that goes close enough to the receiver. If there is lateral deviation then both ϕ and θ must be changed for successive rays. A two-dimensional form of the rule of false position can then be devised. To use it, three rays must first be traced; see problem 14.2.

A method of this kind in which successive trials are made is sometimes called a 'shooting method'. If the equations are complicated it can be very demanding on computer time. For this type of problem, therefore, ray tracing is customarily used only when some simplification is possible. Another use of ray tracing is to study the configuration of a whole family of rays that leave a given source. For example Herring (1980) used it in this way for a source in the magnetosphere, in a study of radio emissions in the frequency range 10 to 200 kHz.

The first part of this chapter up to § 14.8, like ch. 10, deals almost entirely with loss-free media. The ray direction is then also the direction of the time averaged Poynting vector; see § 5.3. The rays are the paths of energy flux. The next part, §§ 14.9–14.12, discusses the effect of losses in the medium and introduces the concepts of complex rays and the bilinear concomitant vector. The last part, §§ 14.13–14.14, deals with reciprocity when ray theory is used.

The purpose of ray tracing is nearly always to find out the features of the wave that arrives at a given receiver from a given transmitter. A knowledge of the ray path traversed is of secondary importance. It might be useful if it was desired to interrupt the signal by placing an obstacle where it would screen the receiver. The right position would be somewhere on the ray path.

Four uses of ray tracing, (a)–(d), were listed at the end of § 10.1. For items (c), (d) it is necessary to evaluate integrals along the ray path. This can be done by formulating further differential equations, and integrating them at the same time as the differential equations for the ray path are integrated; see § 14.7.

14.2. The eikonal function

Ray tracing is often used where there are sharp boundaries between different media. Where a ray crosses such a boundary its direction, in general, changes discontinuously, and there may be a partial reflection. There are no sharp boundaries in the ionosphere and magnetosphere, and in most of this chapter it is assumed that the medium is continuous and slowly varying. This means that, at points where ray theory can be used, solutions of Maxwell's equations exist that are the analogues of the W.K.B. solutions, ch. 7, or generalisations of them. It was shown in §§ 7.3–7.7 that a W.K.B. solution represents a progressive wave with an exponential or 'phase memory' factor and a more slowly varying amplitude factor; see for example (7.26). The basic formulae of ray tracing use only the phase memory part. Study of the amplitude needs an extension of the theory as given in §§ 10.17, 14.8.

When ray tracing is used, therefore, it is implied that at each point in space there is a progressive wave in which the electric and magnetic fields are given by

$$\boldsymbol{E} = \boldsymbol{E}_0 \exp(-\mathrm{i}\mathscr{E}), \quad \boldsymbol{\mathscr{H}} = \boldsymbol{\mathscr{H}}_0 \exp(-\mathrm{i}\mathscr{E}) \tag{14.2}$$

where the exponential is the phase memory term. The function $\mathscr{E}$ is called the 'eikonal function' (Sommerfeld, 1954). It depends on the space coordinates. Now let (14.2) be substituted in Maxwell's equations (2.44), and use the constitutive relation (4.8). Assume that the amplitude factors $\boldsymbol{E}_0$, $\boldsymbol{\mathscr{H}}_0$ vary so slowly that their spatial derivatives may be ignored. Then

$$\boldsymbol{\Gamma}\boldsymbol{E}_0 = \boldsymbol{\mathscr{H}}_0, \quad \boldsymbol{\Gamma}\boldsymbol{\mathscr{H}}_0 = -\boldsymbol{\varepsilon}\boldsymbol{E}_0, \tag{14.3}$$

where $\boldsymbol{\Gamma}$ is the matrix

$$\boldsymbol{\Gamma} = \begin{pmatrix} 0 & -\partial\mathscr{E}/\partial z & \partial\mathscr{E}/\partial y \\ \partial\mathscr{E}/\partial z & 0 & -\partial\mathscr{E}/\partial x \\ -\partial\mathscr{E}/\partial y & \partial\mathscr{E}/\partial x & 0 \end{pmatrix}. \tag{14.4}$$

But (14.3) is the same as (4.70) provided that

$$\boldsymbol{n} = \operatorname{grad}\mathscr{E}. \tag{14.5}$$

The components of $\boldsymbol{n}$ satisfy the dispersion relation in the form (4.73). Thus (14.5) is the condition that (14.2) shall satisfy Maxwell's equations. It is called the 'eikonal equation'. It shows that $\mathscr{E}$ is similar to a potential function in space, and its gradient $\boldsymbol{n}$ is an irrotational vector field, so that

$$\operatorname{curl}\boldsymbol{n} = 0. \tag{14.6}$$

If $\mathscr{E}$ is known at any point x_0, y_0, z_0, its value at any other point x, y, z is found by integrating its gradient along any path from A to B. Thus

$$\mathscr{E} = \int_{x_0}^{x} n_x \, \mathrm{d}x + \int_{y_0}^{y} n_y \, \mathrm{d}y + \int_{z_0}^{z} n_z \, \mathrm{d}z + \text{constant} \tag{14.7}$$

where the same path is used for all three integrals.

An example of a progressive wave for a plane stratified medium was given by (10.4). Here S_{10}, S_{20}, q are the same as the components n_x, n_y, n_z of the refractive index vector $\boldsymbol{n}$. The exponent could be written $\exp(-\mathrm{i}\mathscr{E})$ where $\mathscr{E}$ is given by (14.7). In the plane stratified medium Snell's law applies and n_x, n_y in (10.4) are constants. Thus the first two terms of (14.7) give $xn_x + yn_y + \text{constant}$. The remaining term $\int q\,\mathrm{d}z = \int n_z\,\mathrm{d}z$ was called the phase memory term. It was discussed in § 7.3, and used in (10.2). When n_x, n_y are not constant their integrals in (14.7) must be retained. This shows how the eikonal (14.7) is simply an extension to three dimensions of the phase memory concept.

In general the field in the neighbourhood of any point in space is more complicated than the progressive wave (14.2). It may be the sum of two or more progressive waves. It must then be resolved into its component progressive waves, and each ray is associated with one only of these components. The others are studied separately, or ignored. But each component progressive wave is, by itself, a solution of Maxwell's equations.

14.3. The canonical equations for a ray path

The differential equations for a ray path are now to be found. It will be shown that six variables are needed, namely the Cartesian coordinates x, y, z in ordinary space and the components n_x, n_y, n_z of the refractive index vector $\boldsymbol{n}$, in refractive index space. These six variables may be thought of as defining a six-dimensional space.

Let x, y, z be the coordinates of the point where some feature of the wave intersects the ray, at an instant of time $t = \tau$. This feature could be, for example, a wave crest. Then $\mathrm{d}x/\mathrm{d}\tau$, $\mathrm{d}y/\mathrm{d}\tau$, $\mathrm{d}z/\mathrm{d}\tau$ are the components of the ray velocity $\boldsymbol{V}$ that was studied in ch. 5. Here τ is used as the independent variable, and the total derivative sign $\mathrm{d}/\mathrm{d}\tau$ is used to denote changes that occur on the ray path. Other choices of the independent variable are possible, and this is discussed later. Note particularly that τ is not the time of travel of a wave packet. Now $\boldsymbol{V}$ can be found when the dispersion relation is known.

At any point x, y, z, the dispersion relation may be written

$$D(x, y, z; n_x, n_y, n_z) = \text{constant}. \tag{14.8}$$

Various expressions for this were given in § 4.9, for example (4.62)–(4.66) or (4.73), in which the constant is zero, but other forms are possible. Some authors use the form $\omega(x, y, z; \kappa_x, \kappa_y, \kappa_z) = \omega_0$ where ω_0 is the constant frequency of the wave, and they use the vector $\boldsymbol{\kappa}=\omega\boldsymbol{n}/c$ instead of the vector $\boldsymbol{n}$ used here. See for example, Suchy and Paul (1965), Buckley (1982). This form is avoided here because of its algebraic complexity; see end of § 5.8. Possible choices for D in (14.8) are discussed in §§ 14.4, 14.5. At a fixed point x, y, z (14.8) is the equation of the refractive index surface in refractive index space. For a point on this surface, $\boldsymbol{V}$ has the direction of the normal; see § 5.3. Thus

$$\boldsymbol{V}=(\mathrm{d}x/\mathrm{d}\tau, \mathrm{d}y/\mathrm{d}\tau, \mathrm{d}z/\mathrm{d}\tau)=A(\partial D/\partial n_x, \partial D/\partial n_y, \partial D/\partial n_z). \tag{14.9}$$

The multiplier A may be found from (5.36) with x, y, z instead of ξ, η, ζ. Substitution of (14.9) for $\boldsymbol{V}$ shows that

$$A=c/\{n_x\partial D/\partial n_x+n_y\partial D/\partial n_y+n_z\partial D/\partial n_z\}. \tag{14.10}$$

Equations (14.9) with (14.10) are the required differential equations of the ray path. To use them, however, it is necessary to know the correct values of n_x, n_y, n_z. These change as the point x, y, z moves along the ray, and further equations are needed to determine them.

The equations (14.9) depend on D at one point only in space. They take no cognisance of how the medium varies in space. The additional equations must allow for this variation because it is the spatial gradient of the composition of the medium that determines the curvature of the ray. Thus we have to study how D varies in space, not only for points on the ray path but for adjacent points as well.

Equation (14.8) must hold for all points. Consider an infinitesimal displacement δx from a point x, y, z on the ray, with y and z held constant. Then the components n_x, n_y, n_z change but D remains constant. Hence

$$\frac{\delta D}{\delta x}=0=\frac{\partial D}{\partial x}+\frac{\partial D}{\partial n_x}\frac{\partial n_x}{\partial x}+\frac{\partial D}{\partial n_y}\frac{\partial n_y}{\partial x}+\frac{\partial D}{\partial n_z}\frac{\partial n_z}{\partial x}. \tag{14.11}$$

Because the displacement δx is infinitesimal, the partial derivatives on the right all take their values on the ray. Now (14.6) shows that

$$\frac{\partial n_y}{\partial x}=\frac{\partial n_x}{\partial y}, \quad \frac{\partial n_z}{\partial x}=\frac{\partial n_x}{\partial z}. \tag{14.12}$$

If this and (14.9) are used in (14.11) it gives:

$$A\frac{\partial D}{\partial x}+\frac{\partial n_x}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}\tau}+\frac{\partial n_x}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}\tau}+\frac{\partial n_x}{\partial z}\frac{\mathrm{d}z}{\mathrm{d}\tau}=0. \tag{14.13}$$

The last three terms are simply $\mathrm{d}n_x/\mathrm{d}\tau$. A similar argument applies if $\delta D/\delta y$ or $\delta D/\delta z$ is used in (14.11). Hence we obtain the three equations

$$(\mathrm{d}n_x/\mathrm{d}\tau, \mathrm{d}n_y/\mathrm{d}\tau, \mathrm{d}n_z/\mathrm{d}\tau)=-A(\partial D/\partial x, \partial D/\partial y, \partial D/\partial z). \tag{14.14}$$

The six equations (14.9) and (14.14) are the required differential equations for the ray path. They and the arguments that led to them can be found in the work of Sir William R. Hamilton (collected papers 1931; the original papers were written in

1827–32), and he called them the 'canonical' equations of the ray. But the notation he used is quite different from that used here. The ideas permeate a large part of his papers and it is not easy to give a precise page reference. A study of Hamilton's work is essential and rewarding for anyone who aspires to be a specialist in ray theory, but cannot be recommended to the radio engineer who wishes to acquire a practical working knowledge of ray tracing techniques.

Equations (14.14) are a generalisation of Snell's law. For example suppose they are applied to a horizontally stratified ionosphere as studied in ch. 10. Then the dispersion relation is independent of x and y, so that $\partial D/\partial x$, $\partial D/\partial y$ are both zero. Hence n_x and n_y are constant on a ray, and this is Snell's law, as discussed in § 6.2. They are the same as S_1, S_2 respectively.

For a general medium, (14.14) can be used to define an 'effective' local surface of stratification near any point. The vector $\partial D/\partial x$, $\partial D/\partial y$, $\partial D/\partial z$ is the normal to this surface, and will be called the 'effective stratification vector'. For any plane stratified medium it is normal to the strata and its direction is the same for all possible ray directions as given by (14.9). But in general it is different for different ray directions, and there is no unique surface of stratification. For a counter example see problem 14.1.

14.4. Properties of the canonical equations

The coordinates x, y, z of a variable point on a ray may be written as a vector $\boldsymbol{r}$. The operator grad is then often written grad $D = \partial D/\partial \boldsymbol{r}$. With this notation the canonical equations (14.9) (14.14) in vector form are

$$\mathrm{d}\boldsymbol{r}/\mathrm{d}\tau = A\partial D/\partial \boldsymbol{n}, \quad \mathrm{d}\boldsymbol{n}/\mathrm{d}\tau = -A\partial D/\partial \boldsymbol{r}, \tag{14.15}$$

where from (14.10)

$$A = c/(\boldsymbol{n}\cdot\partial D/\partial \boldsymbol{n}). \tag{14.16}$$

For an isotropic medium the ray and the wave normal have the same direction, and the refractive index n is a scalar function of position. The dispersion relation may be written

$$D \equiv (n_x^2 + n_y^2 + n_z^2)^{\frac{1}{2}} - n(x, y, z) = 0. \tag{14.17}$$

Then (14.16) gives $A = c/n$ and the equations (14.15) are

$$\frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}\tau} = c\frac{\boldsymbol{n}}{n^2}, \quad \frac{\mathrm{d}\boldsymbol{n}}{\mathrm{d}\tau} = \frac{c}{n}\frac{\partial n}{\partial \boldsymbol{r}}. \tag{14.18}$$

For a more general anisotropic medium, when the canonical equations in the form (14.15) are integrated with a step-by-step process they give $\boldsymbol{r}$ and $\boldsymbol{n}$ at successive points along a ray. The value of $\boldsymbol{n}$ ought to satisfy the dispersion relation (14.8), and for an exact solution it would do so because the equations were derived from (14.8). But in numerical integration rounding errors accumulate and lead to a wrong value of $\boldsymbol{n}$ and thence of D and its derivatives. It is advisable, therefore, to use the dispersion relation as the integration proceeds. The rate at which errors accumulate depends on the step size used in the integration. One method that is often used is to carry out the integration for successive groups of steps, and test the dispersion relation at the end of each group. If it is not satisfied with sufficient accuracy, the last group is repeated with a smaller step size.

The form (14.15) is not always the most convenient or simplest and other versions are often used. For example there is one method in which one of the three components of $\mathrm{d}\boldsymbol{n}/\mathrm{d}\tau$ from (14.15) is omitted, and for illustration we shall suppose that it is the equation for $\partial n_z/\partial\tau$. Integration of the remaining two equations gives n_x, n_y, and the dispersion relation is then used to find n_z. There are many other ways of using a modified form of the equations. The choice depends on the particular problem being solved. The methods are best illustrated by describing some examples of particular problems.

Consider the special case of an electron plasma whose composition is independent of y, and let the ray, the wave normal and the vector $\boldsymbol{Y}$ all lie in the plane $y = 0$. Then they remain in this plane and $y = 0$, $n_y = 0$. Only four of the six equations (14.15) are non-trivial. Let χ be the angle between the wave normal and the n_x axis so that

$$\tan\chi = n_z/n_x \tag{14.19}$$

whence

$$\frac{\mathrm{d}\chi}{\mathrm{d}\tau} = \frac{\partial\chi}{\partial n_x}\frac{\mathrm{d}n_x}{\mathrm{d}\tau} + \frac{\partial\chi}{\partial n_z}\frac{\mathrm{d}n_z}{\mathrm{d}\tau} = \frac{1}{n^2}\left(n_x\frac{\mathrm{d}n_z}{\mathrm{d}\tau} - n_z\frac{\mathrm{d}n_x}{\mathrm{d}\tau}\right). \tag{14.20}$$

The refractive index depends on x, z and on the angle χ, and through χ on n_x, n_z. The dispersion relation may be written

$$D \equiv \frac{(n_x^2 + n_z^2)^{\frac{1}{2}}}{n(x, z; \chi)} = 1. \tag{14.21}$$

This form is discussed in more detail in § 14.5. In (14.16) it gives $A = c$. Then the canonical equations (14.15) give, with (14.20)

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}\tau} &= \frac{c}{n}\left(\cos\chi + \sin\chi\frac{1}{n}\frac{\partial n}{\partial\chi}\right),\\ \frac{\mathrm{d}z}{\mathrm{d}\tau} &= \frac{c}{n}\left(\sin\chi - \cos\chi\frac{1}{n}\frac{\partial n}{\partial\chi}\right),\\ \frac{\mathrm{d}\chi}{\mathrm{d}\tau} &= \frac{c}{n^2}\left(\cos\chi\frac{\partial n}{\partial z} - \sin\chi\frac{\partial n}{\partial x}\right). \end{aligned} \tag{14.22}$$

When (14.22) are integrated, the refractive index n can be found at each point, since χ is known. This shows how the two equations for $\mathrm{d}n_x/\mathrm{d}\tau$ and $\mathrm{d}n_z/\mathrm{d}\tau$ may be replaced by a single equation for $\mathrm{d}\chi/\mathrm{d}\tau$ together with the formula for $n(x, z; \chi)$. The equations (14.22) were given by Haselgrove (1954) and were used by her (Haselgrove, 1957) to study rays in the magnetic meridian plane in the ionosphere.

The multiplier A in (14.15) depends on the choice of the function D in the dispersion relation (14.8). Its value does not affect the path of the ray nor the values of $\boldsymbol{n}$ on it. It is often combined with τ to give a new independent variable s, so that equations (14.15) become

$$\mathrm{d}\boldsymbol{r}/\mathrm{d}s = \partial D/\partial \boldsymbol{n}, \quad \mathrm{d}\boldsymbol{n}/\mathrm{d}s = -\partial D/\partial \boldsymbol{r}. \tag{14.23}$$

Suppose now that we wish to trace a ray that goes from free space into the ionosphere, and suppose that (4.62) is chosen for the dispersion relation. It applies for a particular set of axes but in the free space it is the same for all axis systems and reduces to

$$D \equiv n^4 - 2n^2 + 1 = 0. \tag{14.24}$$

Now all elements of $\partial D/\partial \boldsymbol{n}$ are zero, so that A (14.16) is infinite. When ds in (14.23) is finite, dτ must be zero, and d$\boldsymbol{r}$, d$\boldsymbol{n}$ are both zero. We cannot make any progress along the ray. Clearly a different D must be chosen. A possible choice is discussed in the following section.

For a ray that enters the ionosphere from free space there is a further problem. It must be either ordinary or extraordinary, and for both, at the start, $n = 1$. There is nothing in equations (14.15) that will decide which of the two types it is. This may be dealt with by writing, in the dispersion relation,

$$n^2 = 1 + Xp \tag{14.25}$$

and expressing it in terms of p rather than n. Now (4.112) shows that in the limit of free space, $X \to 0$, p tends to a non-zero limit that is different for the ordinary and extraordinary waves. Thus at the start the value of p for the required wave is used. The details are complicated, and are given by Haselgrove and Haselgrove (1960) and by Haselgrove (1963). They use q instead of p but it is avoided here because of possible confusion with Booker's q, ch. 6.

For other versions of the ray tracing equations see, for example, Al'pert (1948), Kimura (1966), Bitoun, Graff and Aubry (1970), Suchy (1972a, b, 1974a, 1981), Rönnmark (1984). See also *Radio Science* (1968, special issue on ray tracing).

If the transmitter or receiver is moving, or if the medium is changing with time, the ray tracing problem is more complicated and the received signal can show a Doppler shift of frequency. The effect is not discussed in this book, but see for example Little and Lawrence (1960), Capon (1961), Bennett (1969).

14.5. The Haselgrove form of the equations

The refractive index n is given by the Appleton–Lassen formula (4.48) or some equivalent form such as (4.51), (4.61). The variables in it are functions of the space coordinates and it also contains the angle Θ, that depends on the direction cosines of the wave normal:

$$\frac{n_x}{(n_x^2+n_y^2+n_z^2)^{\frac{1}{2}}},\quad \frac{n_y}{(n_x^2+n_y^2+n_z^2)^{\frac{1}{2}}},\quad \frac{n_z}{(n_x^2+n_y^2+n_z^2)^{\frac{1}{2}}}. \tag{14.26}$$

Thus n may be written $n(x,y,z;\ n_x,n_y,n_z)$ where n_x,n_y,n_z occur only in the combinations (14.26). Now in place of the function D in the dispersion relation (14.8), we choose a new function G so that the dispersion relation is

$$G(x,y,z;\ n_x,n_y,n_z) \equiv \frac{(n_x^2+n_y^2+n_z^2)^{\frac{1}{2}}}{n(x,y,z;\ n_x,n_y,n_z)} = 1. \tag{14.27}$$

If now n_x,n_y,n_z are replaced by $\lambda n_x, \lambda n_y, \lambda n_z$, the combinations (14.26) are unaltered, and G changes to λG. Therefore G is said to be homogeneous of degree one in the set n_x, n_y, n_z, and Euler's theorem for homogeneous functions shows that

$$n_x\partial G/\partial n_x + n_y\partial G/\partial n_y + n_z\partial G/\partial n_z = G. \tag{14.28}$$

The proof of the theorem is not difficult and is given in standard text books (see, for example, Gibson, 1929, p. 412). Now (14.28) and (14.16) with G for D show that, since $G=1$, $A=c$, and (14.15) becomes:

$$\mathrm{d}\boldsymbol{r}/\mathrm{d}\tau = c\partial G/\partial \boldsymbol{n},\quad \mathrm{d}\boldsymbol{n}/\mathrm{d}\tau = -c\partial G/\partial \boldsymbol{r}. \tag{14.29}$$

Equations of this form were given by Hamilton (1931) who also used the properties of homogeneous functions. They were derived and used for radio propagation problems by Haselgrove (1954, 1957, 1963), Haselgrove and Haselgrove (1960), and they are commonly called the Haselgrove equations. A simple example (14.22) has already been given. When the equations are expressed in Cartesian coordinates, as used in (14.27), various transformations are possible that greatly simplify the algebra (see the Haselgrove references cited above). This is true even when they are applied in an ionosphere or magnetosphere with spherical or more complicated stratification, because the simplicity of the Cartesian equations outweighs the extra complication of expressing a spherical or more irregular plasma in Cartesian coordinates (Haselgrove, 1963).

The equations are, however, often used in spherical polar coordinates r, θ, ϕ, and their basic form for this case can be derived as follows. At each point in space the components of the refractive index vector are n_r parallel to the radius r in ordinary space, n_θ parallel to the direction $r=$ const., $\phi=$ const., and n_ϕ parallel to the direction $r=$ const., $\theta=$ const. Thus the refractive index space is a Cartesian space, but the directions of its axes are different for different points in ordinary space. The dispersion relation is

$$G \equiv \mathscr{G}(r,\theta,\phi;\ n_r,n_\theta,n_\phi) \equiv \frac{(n_r^2+n_\theta^2+n_\phi^2)^{\frac{1}{2}}}{\breve{n}(r,\theta,\phi;\ n_r,n_\theta,n_\phi)} = 1. \tag{14.30}$$

Here $\breve{n}$ is numerically the same as n but a different symbol is needed because the functional dependence on the new variables is different. Similarly $\mathscr{G}$ is used instead of G. The equations for transforming the variables are

$$r = (x^2 + y^2 + z^2)^{\frac{1}{2}}, \quad \tan\theta = (x^2 + y^2)^{\frac{1}{2}}/z, \quad \tan\phi = y/x \tag{14.31}$$

$$\left.\begin{aligned} n_r &= (n_x \cos\phi + n_y \sin\phi)\sin\theta + n_z\cos\theta, \\ n_\theta &= (n_x\cos\phi + n_y\sin\phi)\cos\theta - n_z\sin\theta, \\ n_\phi &= -n_x\sin\phi + n_y\cos\phi. \end{aligned}\right\} \tag{14.32}$$

Hence from (14.29), after some algebra

$$\frac{\mathrm{d}r}{\mathrm{d}\tau} = c\frac{\partial\mathscr{G}}{\partial n_r}, \quad \frac{\mathrm{d}\theta}{\mathrm{d}\tau} = \frac{c}{r}\frac{\partial\mathscr{G}}{\partial n_\theta}, \quad \frac{\mathrm{d}\phi}{\mathrm{d}\tau} = \frac{c}{r\sin\theta}\frac{\partial\mathscr{G}}{\partial n_\phi}, \tag{14.33}$$

$$\left.\begin{aligned} \frac{\mathrm{d}n_r}{\mathrm{d}\tau} &= c\left(-\frac{\partial\mathscr{G}}{\partial r} + \frac{n_\theta}{r}\frac{\partial\mathscr{G}}{\partial n_\theta} + \frac{n_\phi}{r}\frac{\partial\mathscr{G}}{\partial n_\phi}\right), \\ \frac{\mathrm{d}n_\theta}{\mathrm{d}\tau} &= c\left(-\frac{1}{r}\frac{\partial\mathscr{G}}{\partial\theta} - \frac{n_r}{r}\frac{\partial\mathscr{G}}{\partial n_\theta} + \frac{n_\phi\cot\theta}{r}\frac{\partial\mathscr{G}}{\partial n_\phi}\right), \\ \frac{\mathrm{d}n_\phi}{\mathrm{d}\tau} &= c\left(-\frac{1}{r\sin\theta}\frac{\partial\mathscr{G}}{\partial\phi} - \frac{n_r + n_\theta\cot\theta}{r}\frac{\partial\mathscr{G}}{\partial n_\phi}\right). \end{aligned}\right\} \tag{14.34}$$

14.6. Fermat's principle

Fermat's principle is usually stated as follows. Consider several different paths running from a fixed point A to another fixed point B. Suppose that a ray can travel along any one of these paths and let T be the time taken for some feature of the wave, such as a wave crest or trough, to go from A to B. It is assumed that this feature moves with the ray velocity in the direction of the path. Then the actual ray that goes from A to B is that path for which T is an extremum.

For an isotropic medium the proof is not difficult. The ray velocity V at each point is c/n in the direction of the wave normal and is the same for all ray directions. This gives at once the first set of the ray tracing equations (14.18), that is

$$\frac{\mathrm{d}x}{\mathrm{d}\tau} = cn_x/n^2 \tag{14.35}$$

and two similar equations with y, n_y, and z, n_z. Any path from A to B is defined by specifying the three functions $x(p)$, $y(p)$, $z(p)$ where p is some parameter that increases monotonically from 0 at A to 1 at B. It is convenient to use the notation

$$x' = \mathrm{d}x/\mathrm{d}p, \quad y' = \mathrm{d}y/\mathrm{d}p, \quad z' = \mathrm{d}z/\mathrm{d}p. \tag{14.36}$$

Then, if $\mathrm{d}s$ is an element of the path,

$$\frac{\mathrm{d}s}{\mathrm{d}p} = \{x'^2 + y'^2 + z'^2\}^{\frac{1}{2}} = \frac{c\,\mathrm{d}\tau}{n\,\mathrm{d}p}. \tag{14.37}$$

The time of travel from A to B is

$$T = \int_{(A)}^{(B)} \mathrm{d}\tau = \frac{1}{c}\int_0^1 n\frac{\mathrm{d}s}{\mathrm{d}p}\mathrm{d}p. \tag{14.38}$$

If, for some path $\mathscr{P}$, this is an extremum, then on using another path infinitesimally different from $\mathscr{P}$ the change δT of T must be zero. This is a standard problem in the calculus of variations. The necessary and sufficient condition that $\delta T = 0$ is given by the three Euler equations

$$\frac{\partial}{\partial x}\left\{n\frac{\mathrm{d}s}{\mathrm{d}p}\right\} = \frac{\mathrm{d}}{\mathrm{d}p}\left\{\frac{\partial}{\partial x'}\left(n\frac{\mathrm{d}s}{\mathrm{d}p}\right)\right\} \tag{14.39}$$

and two similar equations for the y and z components. On carrying out the differentiations and using (14.36), (14.37), this gives

$$\frac{\partial n}{\partial x}\frac{\mathrm{d}s}{\mathrm{d}p} = \frac{\mathrm{d}}{\mathrm{d}p}\left\{nx'/\frac{\mathrm{d}s}{\mathrm{d}p}\right\} = \frac{\mathrm{d}}{\mathrm{d}p}\left\{n\frac{\mathrm{d}x}{\mathrm{d}s}\right\} = \frac{\mathrm{d}}{\mathrm{d}p}\left\{\frac{n^2}{c}\frac{\mathrm{d}x}{\mathrm{d}\tau}\right\}. \tag{14.40}$$

Then on rearranging and using (14.36)

$$\frac{\partial n}{\partial x} = \frac{\mathrm{d}n_x}{\mathrm{d}s} = \frac{n}{c}\frac{\mathrm{d}n_x}{\mathrm{d}\tau}. \tag{14.41}$$

This and the two similar equations for the y and z components are the same as the second set of the ray tracing equations (14.18). Thus these equations are the Euler equations that are the equivalent of Fermat's principle.

For an anisotropic medium the problem is more complicated. The magnitude of the ray velocity V now depends on its direction. It is given by $V = c/n\cos\alpha$ where $n\cos\alpha$ is the ray refractive index $\mathscr{M}$, § 5.3. Thus

$$n\cos\alpha = \mathscr{M}(x, y, z; V_x, V_y, V_z) \tag{14.42}$$

where the dependence on V_x, V_y, V_z is only through the direction cosines

$$\frac{V_x}{(V_x^2+V_y^2+V_z^2)^{\frac{1}{2}}},\quad \frac{V_y}{(V_x^2+V_y^2+V_z^2)^{\frac{1}{2}}},\quad \frac{V_z}{(V_x^2+V_y^2+V_z^2)^{\frac{1}{2}}} \tag{14.43}$$

of the ray direction. The equation of the ray surface at the point x, y, z may now be written

$$F(x, y, z; V_x, V_y, V_z) \equiv \frac{1}{c}(V_x^2+V_y^2+V_z^2)^{\frac{1}{2}}\mathscr{M}(x, y, z; V_x, V_y, V_z) = 1. \tag{14.44}$$

It is homogeneous of degree 1 in V_x, V_y, V_z and therefore, by Euler's theorem (compare (14.28)),

$$V_x\partial F/\partial V_x + V_y\partial F/\partial V_y + V_z\partial F/\partial V_z = F = 1. \tag{14.45}$$

The function F gives the ray surface and is a kind of analogue of the function G, (14.27), that gives the refractive index surface.

The ray surface has the shape of a wave front, see end of § 5.3, and therefore its normal is the wave normal, parallel to $\boldsymbol{n}$. Thus

$$\boldsymbol{n} = c\partial F/\partial \boldsymbol{V} \tag{14.46}$$

where (5.36) and (14.45) have been used to give the factor c.

Now we apply to F an argument similar to that used for D in (14.11)–(14.13). Let δx be an infinitesimal displacement from a point on the ray, with y and z held constant. Then

$$\frac{\delta F}{\delta x} = 0 = \frac{\partial F}{\partial x} + \frac{1}{c}\left(n_x\frac{\partial V_x}{\partial x} + n_y\frac{\partial V_y}{\partial y} + n_z\frac{\partial V_z}{\partial z}\right) \tag{14.47}$$

where (14.46) has been used. The same argument can be applied to G and it gives

$$\frac{\delta G}{\delta x} = 0 = \frac{\partial G}{\partial x} + \frac{1}{c}\left(V_x\frac{\partial n_x}{\partial x} + V_y\frac{\partial n_y}{\partial x} + V_z\frac{\partial n_z}{\partial x}\right) \tag{14.48}$$

where the first equation (14.29) has been used. Addition of (14.47) and (14.48) now gives

$$\frac{\partial}{\partial x}(F+G) = -\frac{1}{c}\frac{\partial}{\partial x}(\boldsymbol{V}\cdot\boldsymbol{n}) = 0 \tag{14.49}$$

from (5.36) and there are similar results for $\partial/\partial y$ and $\partial/\partial z$. Thence from the second equation (14.29)

$$c\partial F/\partial x = \mathrm{d}n_x/\mathrm{d}\tau \tag{14.50}$$

and there are two similar equations with y and z.

Now we again use $x(p)$, $y(p)$, $z(p)$ to specify paths from A to B, and $x' = \mathrm{d}x/\mathrm{d}p$ etc. as in (14.36). Then we have $x' = V_x\mathrm{d}\tau/\mathrm{d}p$ and two similar equations with y and z. There is some latitude in the choice of the parameter p and we choose $p = \tau/T$ so that, for any one path, $\mathrm{d}\tau/\mathrm{d}p$ is a constant. Then

$$\frac{\mathrm{d}s}{\mathrm{d}p} = \{V_x^2 + V_y^2 + V_z^2\}^{\frac{1}{2}}\frac{\mathrm{d}\tau}{\mathrm{d}p}. \tag{14.51}$$

Thus $\mathrm{d}\tau = (\mathscr{M}/c)(\mathrm{d}s/\mathrm{d}p)\mathrm{d}p$ and this is now used in (14.38). It shows that, for the anisotropic medium, the refractive index n is to be replaced by the ray refractive index $\mathscr{M}$. For $\delta T = 0$ the Euler equations, for (14.38) with $\mathscr{M}$ for n, are now

$$\frac{\partial}{\partial x}\left\{\mathscr{M}\frac{\mathrm{d}s}{\mathrm{d}p}\right\} = \frac{\mathrm{d}}{\mathrm{d}p}\left\{\frac{\partial}{\mathrm{d}x'}\left(\mathscr{M}\frac{\mathrm{d}s}{\mathrm{d}p}\right)\right\} \tag{14.52}$$

and two similar equations; compare (14.39). Cancel the constant factor $\mathrm{d}\tau/\mathrm{d}p$ and use (14.44). Then

$$\frac{\partial F}{\partial x} = \frac{\mathrm{d}}{\mathrm{d}p}\left(\frac{\partial F}{\partial x'}\right) = \frac{\mathrm{d}}{\mathrm{d}\tau}\left(\frac{\partial F}{\partial V_x}\right). \tag{14.53}$$

Now (14.50) and (14.46) show that the first and last terms are both equal to $(1/c)(\mathrm{d}n_x/\mathrm{d}\tau)$. Similarly the other two Euler equations are satisfied. Thus Fermat's principle is established. It is again equivalent to the ray tracing equations (14.29) which were used at the stages (14.48) and (14.50).

Fermat's principle is sometimes asserted as an *a priori* postulate, and used to derive the ray tracing equations. This is not satisfactory, because it is necessary first to formulate the definition of a ray and this involves the restriction to a progressive wave as in § 14.2. Moreover although Maxwell's equations must be satisfied for the fields associated with an allowed ray, they cannot be satisfied for all arbitrary fictitious varied ray paths.

The usefulness of Fermat's principle is that it can be proved for Cartesian coordinates, starting from Maxwell's equations, as has been done here in outline. But it must be independent of the coordinate system and it can therefore then be used to derive the equations in other systems. It was used in this way by Haselgrove (1954) to derive the canonical equations in a very general coordinate system.

One further point should be noted. The time of travel T that is an extremum is not the travel time of a wave packet, but of some feature of the wave itself. This prompts the question: when Fermat's principle is asserted *a priori*, why should the time of travel be that of the wave and not of a wave packet? The wave travel time gives the right answer, but to show this it is necessary to have some other basis for Fermat's principle, so that then it is not an *a priori* postulate at all.

For a discussion of Fermat's principle see, for example, Bennett (1969), Suchy (1972b).

14.7. Equivalent path and absorption

It was shown in § 5.8 that a wave packet travels in the direction of the ray with the group velocity $\mathscr{U}$ given by (5.68), (5.75). The time of travel over a given path from A to B is P'/c where P' is called the equivalent path. Hence

$$P' = c\int_{(\mathrm{A})}^{(\mathrm{B})} \mathrm{d}s/\mathscr{U} = \int_{(\mathrm{A})}^{(\mathrm{B})} n' \cos\alpha \,\mathrm{d}s \tag{14.54}$$

where n' is the group refractive index, §§ 5.8, 5.9 and α is the angle between the ray and the wave normal. Now

$$\mathrm{d}s = V\mathrm{d}\tau = \frac{c\mathrm{d}\tau}{n\cos\alpha} \tag{14.55}$$

whence (14.54) becomes

$$P' = c\int_{(\mathrm{A})}^{(\mathrm{B})} \frac{n'}{n}\mathrm{d}\tau = c\int_{(\mathrm{A})}^{(\mathrm{B})} \left(1 + \frac{f}{n}\frac{\partial n}{\partial f}\right)\mathrm{d}\tau \tag{14.56}$$

from (5.75) first equation, where f is the frequency.

If collisions are allowed for, the refractive index vector $\boldsymbol{n}$ is complex and the wave is attenuated as it travels. The general theory for this case is discussed in § 14.9ff. If the collision frequency is small enough, however, $-\,\mathrm{Im}(\boldsymbol{n})$ is so small that its effect on the ray paths is negligible. The theory of the earlier sections can be applied to Re($\boldsymbol{n}$), and the ray path can be found by ignoring the imaginary part. Once the path is known, the attenuation can be found by an integration along the ray path, as described below.

As we proceed along a ray the signal amplitude may change for two reasons. First, the neighbouring rays in a ray pencil may diverge or converge, so that the energy flux decreases or increases respectively. This was studied for a stratified medium in § 10.17, and it is discussed for the general case in the following section. Second, the signal may be attenuated. The signal amplitude $\mathscr{A}$ is here taken to be the square root of the magnitude of the time averaged Poynting vector (2.63). The following discussion applies only to the attenuation, and not to the divergence or convergence.

Let

$$\boldsymbol{n} = \boldsymbol{\mu} - \mathrm{i}\boldsymbol{\chi}, \quad n = \mu - \mathrm{i}\chi. \tag{14.57}$$

Then χ determines the rate of attenuation in the direction of the wave normal. The attenuation along the ray path is therefore given by

$$\mathscr{A} = \mathscr{A}_0 \exp\left\{ -k \int_{(\mathrm{A})}^{(\mathrm{B})} \chi \cos\alpha \, \mathrm{d}s \right\} \tag{14.58}$$

where $\mathscr{A}_0$ is the signal amplitude at the point A. It is assumed that $\chi \ll \mu$ so that χ can be neglected everywhere except in (14.58). To express $\mathscr{A}$ in terms of the electric field $\boldsymbol{E}$ of the wave, it is assumed that the fields near a point on the ray are given by the progressive wave (14.2) with (14.8). Then it can be shown, from (2.63) and (2.44) first equation that

$$\mathscr{A}^2 = \frac{1}{2Z_0} |\boldsymbol{\mu}(\boldsymbol{E}\cdot\boldsymbol{E}^*) - \mathrm{Re}\{\boldsymbol{E}^*(\boldsymbol{\mu}\cdot\boldsymbol{E})\}|. \tag{14.59}$$

Now (14.55) is replaced by

$$\mathrm{d}s \cos\alpha = c\,\mathrm{d}\tau/\mu \tag{14.60}$$

so that (14.58) gives

$$\ln \mathscr{A} = -kc \int_{(\mathrm{A})}^{(\mathrm{B})} \frac{\chi}{\mu} \mathrm{d}\tau + \ln \mathscr{A}_0. \tag{14.61}$$

The six canonical differential equations, such as (14.29), or their equivalent, are used to find the ray path in a step-by-step integration process. At the same time it is convenient to find the equivalent path and the attenuation by expressing (14.56) and (14.61) as additional differential equations, thus

$$\frac{\mathrm{d}P'}{\mathrm{d}\tau} = c\left(1 + \frac{f}{\mu}\frac{\partial\mu}{\partial f}\right), \tag{14.62}$$

$$\frac{\mathrm{d}(\ln \mathscr{A})}{\mathrm{d}\tau} = -kc\frac{\chi}{\mu}. \tag{14.63}$$

The six canonical equations and the two equations (14.62), (14.63) can then be integrated at the same time.

A slightly different method of ray tracing in an anisotropic absorbing medium was used by Suchy (1972a). A wave packet was launched whose shape was initially a Gaussian function in all three directions. As it travelled in real space it was attenuated and it spread, but the point of maximum signal amplitude could be found at successive times. The locus of this point was taken as the effective real ray and its speed was the generalised group velocity, which was shown to be $\mathrm{Re}(\partial\omega/\partial\boldsymbol{\kappa})$ where $\boldsymbol{\kappa} = \omega\boldsymbol{n}/c$. In later papers (Suchy, 1972b, 1974a) these results were related to a form of ray tracing equations of Hamilton's type in real space.

14.8. Signal intensity in ray pencils

One ray does not exist in isolation. It must be one of a family of neighbouring rays, and we can select a set of rays in this family that lie in the surface of a narrow tube. The rays within this tube form a pencil of rays, and the signal intensity depends on the convergence or divergence of the pencil. But the canonical equations apply to one ray only, and give no information about the neighbouring rays. For this some further equations are needed. The problem was discussed in § 10.17 for rays in a stratified medium. It is now extended for the general medium. In this section it is assumed that the medium is loss-free so that the refractive indices and other associated quantities are all real. The method given here is similar to that of Buckley (1982) but there are some differences. Buckley gives references to similar methods used for acoustic waves, where the medium is isotropic and non-dispersive. The author is indebted to Dr. R. Buckley for permission to use his results, and for discussions. The problem for anisotropic dispersive media has also been studied by Harvey (1968) and Bernstein (1975).

In this section we shall need to use some tensor relations, and therefore vectors and tensors are expressed in Cartesian coordinates with the subscript notation. Thus the vector $\boldsymbol{n}$ is now written n_i, where the subscript i takes one of the three values x, y, z, so that the n_i are the components of $\boldsymbol{n}$. The coordinates x, y, z of a point on a ray are the components of the vector $\boldsymbol{r}$ of § 14.4 and are similarly written r_i. In this notation the Haselgrove form (14.29) of the canonical equations is

$$\mathrm{d}r_i/\mathrm{d}\tau = c\,\partial G/\partial n_i, \quad \mathrm{d}n_i/\mathrm{d}\tau = -c\,\partial G/\partial r_i. \tag{14.64}$$

A quantity can have two or three or more subscripts and is then a tensor of rank two, three, etc. When the same subscript appears on two factors in any product of terms, it is implied that that product is the sum of the three products in which the repeated subscript takes each of the three values x, y, z. This is the summation convention and is fully described by Jeffreys (1931) who also gives the definitions and properties of the isotropic tensors δ_{ij} and ε_{ijk} used below.

It is assumed that a ray pencil originates from a point source where $r_i = 0$. We consider a ray pencil with the property that its cross section by any plane is a small parallelogram. The pencil is then defined by the four rays at the corners. These rays can be specified by giving the four values of the refractive index n_i at the source. They will be indicated by a capital N, and are taken to be

$$N_i, \quad N_i + \delta N_i^{(1)}, \quad N_i + \delta N_i^{(2)}, \quad N_i + \delta N_i^{(1)} + \delta N_i^{(2)}. \tag{14.65}$$

Here $\delta N_i^{(1)}$, $\delta N_i^{(2)}$ play the same role as δS_1, δS_2 in § 10.17. The ray specified by N_i is the 'main' ray, to be traced by using the canonical equations. The other rays are infinitesimally displaced from it. Let the points where these four rays cut any given plane be

$$r_i, \quad \delta r_i + r_i^{(1)}, \quad r_i + \delta r_i^{(2)}, \quad r_i + \delta r_i^{(1)} + \delta r_i^{(2)}. \tag{14.66}$$

Then the area of the parallelogram where the pencil crosses the plane is a vector A_i in the direction of the normal and given by the vector product of $\delta r_i^{(1)}$ and $\delta r_i^{(2)}$, thus

$$A_i = \varepsilon_{ijk} \delta r_j^{(1)} \delta r_k^{(2)} \tag{14.67}$$

where ε_{ijk} is the isotropic tensor of rank 3. The energy flux at a point in the pencil is the time averaged Poynting vector $\mathbf{\Pi}_{\mathrm{av}}$ which is now written Π_i. The power crossing the parallelogram is then the scalar product $\Pi_i A_i$. This is the power flux in the pencil and is the same for all cross sections, so that

$$\varepsilon_{ijk} \Pi_i \delta r_j^{(1)} \delta r_k^{(2)} = \text{constant}. \tag{14.68}$$

The signal intensity is the magnitude of Π_i and is here written Π, without subscript as in § 10.17. Thus $\Pi^2 = \Pi_i \Pi_i$. Equation (14.68) is now to be used to find how Π varies along the pencil.

When N_i is changed by δN_i, the values of r_i, n_i at some point on the main ray change to new values for a neighbouring ray. For this type of change the operator $\mathrm{D}/\mathrm{D}N_i$ will be used, to distinguish it from changes along the same ray, for which $\mathrm{d}/\mathrm{d}\tau$ is used. Then

$$\delta r_i = \frac{\mathrm{D} r_i}{\mathrm{D} N_j} \delta N_j, \quad \delta n_i = \frac{\mathrm{D} n_i}{\mathrm{D} N_j} \delta N_j. \tag{14.69}$$

The two 3×3 tensors

$$J_{ij} = \frac{\mathrm{D} r_i}{\mathrm{D} N_j}, \quad K_{ij} = \frac{\mathrm{D} n_i}{\mathrm{D} N_j} \tag{14.70}$$

must be known at any point where (14.69) is to be used. They are found by formulating differential equations for them. Apply the operator $\mathrm{D}/\mathrm{D}N_j$ to (14.64). This gives

$$\frac{\mathrm{d} J_{ij}}{\mathrm{d}\tau} = c \left\{ \frac{\partial^2 G}{\partial n_i \partial r_k} J_{kj} + \frac{\partial^2 G}{\partial n_i \partial n_k} K_{kj} \right\}, \tag{14.71}$$

$$\frac{\mathrm{d} K_{ij}}{\mathrm{d}\tau} = -c \left\{ \frac{\partial^2 G}{\partial r_i \partial r_k} J_{kj} + \frac{\partial^2 G}{\partial r_i \partial n_k} K_{kj} \right\}. \tag{14.72}$$

The tensors (14.70) together have 18 elements and so in (14.71), (14.72) there are 18 simultaneous equations. These are integrated at the same time as the canonical equations (14.64), proceeding along the main ray. The function $G(r_i, n_j)$ is known at all points in ordinary space and refractive index space, so its second partial derivatives used in (14.71) (14.72) are known. At the start of the integration, $r_i = 0$ so all elements of J_{ij} are zero, and $n_i = N_i$ so $K_{ij} = \delta_{ij}$. Thus the initial values are known.

Now the first equation (14.69) is used in (14.68) to give

$$\varepsilon_{ijk}\Pi_i J_{jl} J_{km} \delta N_l^{(1)} \delta N_m^{(2)} = \text{const.} \tag{14.73}$$

The K_{ij} are not used here. They are needed only as auxiliary variables in the equations (14.71), (4.72).

We are free to choose the two constant vectors $\delta N_l^{(1)}$, $\mathrm{d}N_m^{(2)}$ in any way we please, except that they must lie in the refractive index surface at the source point. Thus they must be perpendicular to the normal to this surface, that is to $\partial G/\partial N_i$. A slightly modified form of the choice made by Buckley (1982) is used here:

$$\delta N_l^{(1)} = \left(\frac{\partial G}{\partial N_z}, 0, -\frac{\partial G}{\partial N_x}\right)\delta a^{(1)}, \quad \delta N_m^{(2)} = \left(0, \frac{\partial G}{\partial N_z}, \frac{-\partial G}{\partial N_y}\right)\delta a^{(2)}. \tag{14.74}$$

The product $\delta a^{(1)}\delta a^{(2)}$ of the two scalar amplitude factors appears simply as a constant multiplier on the left side of (14.73) and may be omitted. Then the scalar constant (14.73) is ε_{ijk} multiplied by the three vectors

$$\Pi_i, \quad J_{jx}\frac{\partial G}{\partial N_z} - J_{jz}\frac{\partial G}{\partial N_x}, \quad \text{and } J_{ky}\frac{\partial G}{\partial N_z} - J_{kz}\frac{\partial G}{\partial N_y}. \tag{14.75}$$

It is therefore the 3×3 determinant formed by the components of these vectors (Jeffreys, 1931, p. 13). After rearrangement, and cancellation of the constant factor $\partial G/\partial N_z$, it is more concisely written as the 4×4 determinant

$$\Pi \begin{vmatrix} 0 & \partial G/\partial N_x & \partial G/\partial N_y & \partial G/\partial N_z \\ \hat{\Pi}_x & J_{xx} & J_{xy} & J_{xz} \\ \hat{\Pi}_y & J_{yx} & J_{yy} & J_{yz} \\ \hat{\Pi}_z & J_{zx} & J_{zy} & J_{zz} \end{vmatrix} = \text{constant.} \tag{14.76}$$

The vector Π_i has here been written $\Pi\hat{\Pi}_i$ where $\hat{\Pi}_i$ is a unit vector in the direction of the ray, which is known, and Π is the signal intensity.

Equations (14.76), (14.71), (14.72) are the required solution. To use them, the equations (14.71), (14.72) are integrated at the same time as the canonical equations (14.64), along the main ray. Then at any point where Π is required, the determinant in (14.76) is evaluated. The signal intensity is proportional to its reciprocal.

The signal intensity would be infinite where the rays touch a caustic surface, for there the cross-sectional area of the pencil is zero. Then the determinant in (14.76) is zero. The analogous result to this for a stratified medium was discussed in § 10.17; see equation (10.50). But for a ray that touches a caustic there is no failure of the equations. The canonical equations (14.64) and the set (14.71), (14.72) can be integrated right through a caustic region.

The source of a ray pencil is often in free space or in some homogeneous isotropic medium. It is useful to apply the equations to this case as a simple partial check. Then G is given by (14.27) in which n is a constant, and $\partial G/\partial N_i$ is n_i/n^2. The only non-zero coefficient on the right of (14.71), (14.72) is

$$\frac{\partial^2 G}{\partial n_i \partial n_k} = \frac{1}{n^4}(n^2\delta_{ik} - n_i n_k). \tag{14.77}$$

Equation (14.72) shows that K_{ij} does not change, but retains its initial value δ_{ij}. Then (14.71) can be integrated. Since J_{ij} is initially zero, it gives

$$J_{ij} = \frac{c\tau}{n^4}(n^2\delta_{ij} - n_i n_j). \tag{14.78}$$

The vector $\hat{\Pi}_i$ is n_i/n. These results can be put into (14.76) which gives finally

$$-\Pi c^2\tau^2/n^5 = \text{constant}. \tag{14.79}$$

This shows, correctly, that Π is inversely proportional to R^2 where $R = c\tau/n$ is the distance from the source. The constant is $-1/n^3$ times the power in unit solid angle of the pencil.

For an inhomogeneous isotropic medium the equations (14.76), (14.71), (14.72) take a simpler form, and Buckley (1982) has discussed this case as an illustration. As far as the author knows, at the time of writing no results for an anisotropic medium have been published in which equations of the type (14.71), (14.72) or similar have been used to give computed values of the signal intensity.

14.9. Complex rays. A simple example

The rest of this chapter deals with rays in media for which the losses may be appreciable so that the refractive index $\boldsymbol{n}$ is complex. A method of dealing with this problem when the losses are very small was described in § 10.15 and in § 14.7, equations (14.57)–(14.63). It can be used if $-\mathrm{Im}(\boldsymbol{n})$ is small enough for its effect on the ray path to be neglected, but not for the general case of a lossy medium. The effect of a complex $\boldsymbol{n}$ is that the wave normals, and thence the rays, are refracted so as to have complex directions. Thus it is necessary to consider complex values of the space coordinates x, y, z. The rays must be treated as paths in this complex space. It often happens that a ray path between two real points P and Q is in complex space for almost the whole of its length, and the only points on it that are in real space are the end points P and Q. This idea for radio waves was put forward by Budden and Jull (1964) and a fuller account was given by Jones, R.M. (1970) and by Budden and Terry (1971), and Bennett (1974).

Several other treatments of complex rays have been given. For example Deschamps (1972) introduced the idea of a point source with complex coordinates, and he also studied the energy flux in a ray pencil. Connor and Felsen (1974) were concerned with pulsed waves, or wave packets, and gave examples where it is useful to allow not only the space coordinates but also the time t to take complex values.

When a ray is to be traced through an attenuating medium so as to reach a given receiving point, it is usually necessary to use a 'shooting method', as in § 14.1. Then rays must be traced for successive trial values of the ray direction θ, ϕ at the transmitter. The need to compute with complex variables means that the computing time for any one trial ray is nearly doubled. Since θ and ϕ are now in general complex, their real and imaginary parts comprise four adjustable parameters of the ray, instead of two for real rays. Thus the number of trial complex rays that must be traced is also nearly doubled. The tracing of complex rays has been used for some practical problems, for example radio windows (Budden and Terry, 1971; Budden, 1980; see § 17.8), and ion cyclotron whistlers (Terry 1978; Budden 1983b; see § 13.9). There are ways of partially overcoming some of the complications, but for most practical problems the method is too laborious, and approximate real rays must be used. The importance of complex ray tracing is that the solution it gives is the same as that obtained when the technique of the 'angular spectrum of plane waves' is used, §§ 6.2, 10.1–10.4, and the integrals are evaluated by the method of steepest descents, ch. 9 and §§ 10.2–10.4. For examples see item (7) below and § 14.11. The theory in the earlier sections was given in the language of real variables but applies equally well when the coordinates and the refractive index components S_1, S_2, q are complex. Thus the complex ray provides a 'correct' solution that can be used for testing the shorter approximate methods.

Some workers seem to have difficulty in accepting the idea that a ray can go from P to Q without passing through intervening real points. It may therefore be useful first to illustrate the subject of complex rays by a very simple example.

Consider two semi-infinite homogeneous media separated by a plane boundary $z = 0$. The medium where $z < 0$ is free space and it will be called 'medium 1'. The other medium is isotropic with refractive index n and it will be called 'medium 2'. It is required to trace a ray from the source point P, coordinates $x = y = 0, z = -h_1$, to the receiving point Q, coordinates $x = x_2$, $y = 0$, $z = h_2$. See fig. 14.1. In each medium the ray is straight and in the direction of the wave normal, and must lie in the plane $y = 0$. It is in two segments PC and CQ, that make angles θ_1, θ_2 respectively with the z axis. Snell's law gives

$$\sin\theta_1 = n \sin\theta_2 \tag{14.80}$$

and the geometry of the ray path requires that

$$h_1 \tan\theta_1 + h_2 \tan\theta_2 = x_2. \tag{14.81}$$

The wave is partially reflected at the boundary but the reflected wave is here ignored. Equations (14.80), (14.81) can be converted to an equation of degree 4 in $\tan\theta_1$, but it can be shown that only two of the four solutions are physically acceptable. This subject was studied by Stott (1981) but is not considered further here. If n is real, only one of the physically acceptable solutions has real θ_1 and only this solution will be studied. It corresponds to the ray that would be expected from elementary geometrical optics. Now if n is complex it is easy to see that (14.80), (14.81) can only be satisfied if both θ_1 and θ_2 take complex values. The complex lines PC and CQ in the two media are parts of the two segments of the complex ray from P to Q.

For a ray with given θ_1 in medium 1 the equation of the ray path is

$$x = (z + h_1)\tan\theta_1. \tag{14.82}$$

The boundary is where $z = 0$ and here the ray is refracted. The equation of the ray path in medium 2 is

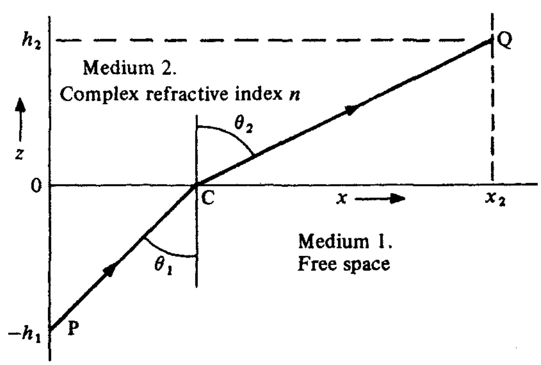
$$x = h_1 \tan\theta_1 + z\tan\theta_2. \tag{14.83}$$

The equation $y = 0$ is also part of the equation for the ray in both media. In the following discussion it is assumed that always $y = 0$ and the problem is treated as a two-dimensional problem in x and z.

For complex n the two equations (14.82), (14.83) are about the simplest possible example of a complex ray. They exhibit most of the important properties, which will now be summarised. The following list gives examples of the locus of a point that moves along a ray starting from the source point P. In all these examples n is complex.

(1) We start from P, that is $x = 0$, $z = -h_1$ along a ray with real θ_1. The equation is (14.82). If z is kept real, then x is real and the ray is a real ray in the usually understood sense. Let z increase until the value $z = 0$ is reached, that is the point C, fig. 14.1. Here the ray is refracted, and in medium 2 its equation is (14.83). But

Fig. 14.1. Schematic representation of a complex ray PCQ that goes between the two real points P, Q. The point C on the boundary is actually at a complex value of x.



(14.80) shows that θ_2 is complex. If we keep z real and non-zero, x is always complex. This ray can never go through the real point Q.

(2) Start again from P with the same real θ_1, that is on the same ray as in (1), and let z change from its initial value $-h_1$ but let it take a complex value. The equation of the ray is still (14.82) which shows that x also is complex. The point (x, z) is in complex space but it is on the same ray. Equation (14.82) expresses x as a function of z in the usual sense where x and z are complex variables. The function $x(z)$ is represented by a Riemann surface in the four-dimensional space of $\mathrm{Re}(x)$, $\mathrm{Im}(x)$, $\mathrm{Re}(z)$, $\mathrm{Im}(z)$. In this example it is a simple plane surface. The real ray of (1), with x and z both real, is a line that is a special cross section of this Riemann surface. In the full sense of the mathematics of complex variables the ray (14.82) is a surface and not a line.

(3) In this example the boundary between the two media is where $z = 0$, and on it z cannot take any other value, real or complex. But x and y can take any values, real or complex. Here we study only the plane $y = 0$. The existence of the boundary does not impose any restriction on x or y.

(4) If the ray in medium 2 is to reach the real point Q, a different ray from that in (1) must be used. Now θ_1 must be complex. If z then increases starting from $-h_1$, and if z is real, x is necessarily complex. When z reaches $z = 0$ the point (x, z) is on the boundary and x has the complex value $h_1 \tan\theta_1$. The complex point $(h_1 \tan\theta_1, 0)$ is the only point that is both on the boundary and on the ray.

(5) Let θ_1 have the same complex value as in (4). Let x increase from zero and remain real. Then z has the complex value $-h_1 + x \cot\theta_1$. There is no real x, except zero, that makes z have a real value. The boundary plane $z = 0$ can never be reached with real x. It is essential to allow x to take complex values.

(6) If we trace the ray by moving along it in small steps of length δs, we have in medium 1

$$\delta x = \delta s \sin\theta_1, \quad \delta z = \delta s \cos\theta_1. \tag{14.84}$$

This ensures that the point remains on the ray (14.82). Now δs is not restricted to real values. We can choose a complex δs to meet any required condition. Since the boundary is at a real value $z = 0$, a natural choice here would be to arrange that z is real so that

$$\arg(\delta s) = -\arg(\cos\theta_1). \tag{14.85}$$

The choice of arg δs is a matter of convenience.

(7) The field at Q from a source at P can be found without recourse to ray theory. The field radiated from P is expressed as an angular spectrum of plane waves. This was done in (10.1) which was for a source at the origin. For a source at P it

becomes

$$F(x,y,z)=\iint A(S_1,S_2)\exp[-ik\{S_1x+S_2y+C(z+h_1)\}]\mathrm{d}S_1\mathrm{d}S_2. \quad (14.86)$$

This gives some field component F in medium 1. The time factor $\exp(ikct)$ is omitted. Each plane wave in the integrand is refracted at the boundary, and Snell's law (14.80) applies to it. Thus the field in medium 2 is

$$F(x,y,z)=\iint B(S_1S_2)\exp\{-ik(S_1x+S_2y+Ch_1+qz)\}\mathrm{d}S_1\mathrm{d}S_2 \quad (14.87)$$

where q is given by (10.12), and $B=TA$. Here $T(S_1,S_2)$ is the transmission coefficient, for example (11.56) or (11.58), at the boundary. Now (14.87) is to be evaluated at the receiving point Q that is $(x_2,0,h_2)$. The method of double steepest descents §§9.10, 10.2 is used. It is assumed that $B=T(S_1,S_2)\ A(S_1,S_2)$ is slowly varying so that the double saddle point is determined from the exponential only. Thus $\partial/\partial S_2$ and $\partial/\partial S_1$ operating on the exponent must give zero. This leads to

$$S_2=0, \quad x_2-h_1S_1/C-h_2S_1(n^2-S_1^2)^{-\frac{1}{2}}=0. \quad (14.88)$$

If now we let

$$S_1=\sin\theta_1, \quad S_1/n=\sin\theta_2 \quad (14.89)$$

then (14.88) gives (14.81). The predominant wave at Q contains the factor

$$\exp\{-ik(x_2\sin\theta_1+h_1\cos\theta_1+nh_2\cos\theta_2)\}. \quad (14.90)$$

But this is exactly the expression that is obtained when the values of θ_1 and θ_2 for the complex ray are used. The complex phase path from P to Q (§ 10.16 and (10.39)) is

$$P=x_2\sin\theta_1+h_1\cos\theta_1+nh_2\cos\theta_2. \quad (14.91)$$

This also, when multiplied by k, is the change of complex phase in going from P to Q along any path in the complex ray.

The component plane waves that make up the integrals (14.86) in medium 1 and (14.87) in medium 2 can be thought of as going from P to Q in real space. The ray is the locus of points for which these integrands have double saddle points when (14.89) is satisfied; see § 10.2 after (10.4). This locus is the complex ray, (14.82), (14.83).

(8) The complex ray that goes from P to Q does not go through any intermediate real point. Suppose we select any real point R in medium 2 somewhere between P and Q. There is a complex ray that goes from P to R, but it is a different ray from the one that goes to Q, and it has different complex values of θ and ϕ

(9) Suppose that, instead of being at P, the source is at infinity in the x–z plane, in a direction that makes a real angle Θ with the z axis. Then the wave in medium

1 is a plane wave. The rays are all parallel to the line from the origin to the source, so any one ray has the equation

$$x = z \tan\Theta + x_0. \tag{14.92}$$

Thus for all these rays $\theta_1 = \Theta$ is real. Here x_0 is the value of x at the boundary $z = 0$. It can take any real or complex value, but rays with differing x_0 are different rays; compare (4) above. If after refraction into medium 2 the ray is to go through Q it is necessary that

$$x_0 + h_2 \tan\theta_2 = x_2. \tag{14.93}$$

But $\sin\theta_2 = \sin\theta_1/n$. Since θ_1 is real, θ_2 is complex. Thus x_0 must be complex. The ray in medium 1 is not in real space but it is still a normal to the incident plane wave.

(10) For more general problems, where the ray is not confined to a plane, it is necessary to let all three coordinates and all three components of n take complex values.

14.10. Real pseudo rays

When losses are very small it is possible at first to ignore $\mathrm{Im}(n)$ and use $\mathrm{Re}(n)$ to trace real rays. The attenuation is then found from the imaginary part of the phase path P that is $\mathrm{Im}(P)$, by integrating $\mathrm{Im}(n)$ along the real ray. This method was described in §§ 10.15, 10.16, 14.7. The question may now be asked: can we use real rays in cases where $\mathrm{Im}(n)$ is not very small?

This question may be examined by using real rays in the simple problem of § 14.9. We therefore suppose that there are rays PC, CQ in fig. 14.1 for which θ_1 and θ_2 have the real values θ_{p1}, θ_{p2}. These rays cannot obey Snell's law (14.80) because n is complex. They will therefore be called 'pseudo rays'. But they can be made to obey a modified form

$$\sin\theta_{p1} = \sin\theta_{p2}\,\mathrm{Re}(n). \tag{14.94}$$

This and (14.81) give real values of θ_{p1}, θ_{p2}. It can be argued that plane waves with normals in the real directions θ_{p1}, θ_{p2} are present in the integrands of (14.86), (14.87) respectively. For these pseudo rays the phase path is

$$P_p = h_1 \sec\theta_{p1} + n h_2 \sec\theta_{p2}. \tag{14.95}$$

Table 14.1 shows some results for the case where medium 2 is a cold isotropic electron plasma with $h_1 = h_2 = \frac{1}{2}x_2$. The values of θ_1, θ_2 were calculated from (14.80), (14.81), P from (14.91), and P_p from (14.95). The real and imaginary parts of all angles are in degrees.

These examples suggest that when $-\arg(n)$ is less than about 10°, the results with real pseudo rays are fairly reliable. For rays in a continuously varying ionosphere this result was confirmed by Budden and Terry (1971) who showed, however,

Table 14.1

X	Z	n	θ_1	θ_2	θ_{p1}	θ_{p2}	P/x_2	P_p/x_2
0.5	0.1	0.711 − 0.035i	34.4 − 3.5i	52.9 + 1.8i	34.5	52.7	1.196 − 0.029i	1.194 − 0.027i
0.5	0.5	0.785 − 0.127i	37.9 − 5.2i	51.2 + 3.3i	35.5	52.2	1.257 − 0.101i	1.213 − 0.014i
10.0	30.0	1.008 − 0.165i	45.5 − 3.1i	44.8 + 3.2i	45.2	44.7	1.425 − 0.116i	1.420 − 0.116i
30.0	30.0	1.086 − 0.460i	53.0 − 8.7i	39.9 + 14.5i	47.2	42.6	1.505 − 0.302i	1.473 − 0.312i
30.0	10.0	1.370 − 1.084i	60.9 − 3.1i	13.5 + 12.4i	52.3	35.3	1.717 − 0.614i	1.656 − 0.664i

that real pseudo rays can give more serious errors in conditions near to penetration of an ionospheric layer. For the last three entries in the table the large values of X and Z are typical for waves of very low frequency, say 10 to 50 kHz in the lower ionosphere. For the last two entries $-\arg(n)$ exceeds 20° and the errors in P_p are more serious. At these frequencies, however, the wavelength is very long and ray theory is rarely used for the ionosphere. It is sometimes used for the magnetosphere, but then Z is much smaller than the values used in the last three entries of the table.

We conclude that the results with real pseudo rays are remarkably good even when $-\arg(n)$ is large, but still it is unwise to use them in cases where there is any doubt, without first checking their accuracy.

14.11. Complex rays in stratified isotropic media

The equations for ray tracing in a stratified ionosphere were derived in §§ 10.2–10.4 and discussed, for an isotropic ionosphere, in § 12.7. Although the discussion was almost entirely in terms of loss-free media and real rays, the equations are still valid when the coordinates and the variables such as $n, q, S = \sin\theta$, take complex values. As an illustration, consider the example used in (12.46). The coordinates are x, y, z with z vertical. The earth's curvature is neglected. There is free space where $0 \leqslant z \leqslant h_0$ and above this $N(z)$ is proportional to $z - h_0$. The plasma is a cold isotropic electron plasma and we now suppose that the collision frequency ν, and thence Z, is independent of height. Then the constant α in (12.46) has a factor $1/(1-\mathrm{i}Z)$, and is therefore complex. Consider a ray that leaves the transmitter at the origin and returns to a receiver on the ground where $x = D$, $y = 0$, $z = 0$. Thus D is the horizontal range, given by (12.48), and must be real. The equation shows that, for this ray, θ must be complex because α is complex. The first and last parts of the ray are in free space and are straight. They must both go through the plane at the base of the ionosphere where $z = h_0$ is real. Here x takes the complex values $h_0 \tan\theta$ and D–$h_0 \tan\theta$ respectively; compare § 14.9 item (4). The ray path within the ionosphere is given by (12.47). It is a relation between the complex variables x, z and represents a Riemann surface. Within the ionosphere it has no solutions for which x and z are both real.

Some numerical results for complex rays in this and other models of the ionosphere were given by Budden and Terry (1971). They compared them with results obtained with real pseudo rays; see end of § 14.10.

A further application of complex rays can be illustrated by using the same model (12.46) of the ionosphere but now with $Z = 0$, so that α is real. This was used in § 12.7 to study real rays. Equation (12.48) gives the horizontal range D to a receiver on the ground, and it can be shown that the phase path is

$$P = 2h_0 \sec\theta + \frac{2f^2}{\alpha}(\sin^2\theta + \tfrac{1}{3}\cos^2\theta)\cos\theta. \tag{14.96}$$

Now (12.48) can be converted to a cubic equation for $\tan\theta$ with real coefficients. It always has one real solution and it may have three. This was illustrated in fig. 12.9. Consider a value of D that is small enough for the cubic to have only one real solution. Then the receiver is inside a skip distance. The other two solutions of the cubic give complex θ and correspond to complex rays. For these two θs the imaginary parts have opposite signs and it can be shown that the values of $\mathrm{Im}(P)$ also have opposite signs. On the complex ray for which $\mathrm{Im}(P)$ is negative, the signal is attenuated. It is present but with small amplitude. For the complex ray with positive $\mathrm{Im}(P)$ the signal would be enhanced, but this is not possible on energy grounds. This ray therefore cannot be present.

The skip distance is where a caustic surface crosses the ground. Near it the signal amplitude is given by an expression containing an Airy integral function; see (10.67) and § 10.22. On the dark side of the caustic, that is within the skip distance, the asymptotic approximation for the Airy integral function has only a subdominant term that decreases as we move further away from the caustic, and this is the signal from the complex ray with negative $\mathrm{Im}(P)$. The other complex ray with positive $\mathrm{Im}(P)$ would correspond to the absent dominant term. These conclusions can be confirmed, without recourse to ray theory, by using the method of the angular spectrum of plane waves, § 10.2. They were illustrated by Budden and Terry (1971). See also Connor and Felsen (1974).

14.12. Complex rays in anisotropic absorbing media

In an anisotropic medium with losses, the refractive index surfaces are complex surfaces in complex refractive index space. It was shown in § 5.3 by two different methods that the ray has the direction of the normal to the refractive index surface. In the second method, equations (5.29)–(5.31), the principle of stationary phase in an angular spectrum of plane waves was used. It can still be applied for complex $\boldsymbol{n}$, and the condition (5.31) still holds. Here ξ, η, ζ is a vector $\boldsymbol{g}$ parallel to the ray. The phase (5.30) now takes complex values. In general, for a given $\boldsymbol{n}$, the direction of $\boldsymbol{g}$ in (5.31) is complex. The formulae (5.33)–(5.35) for V, v, α are still true and all these quantities are now complex.

The equation of the refractive index surface is the dispersion relation. It may be written in the form (14.8) where (x, y, z) is a point in ordinary space and may be complex. Then the ray $\boldsymbol{g}$ is parallel to the normal, that is to the right-hand expression in (14.9), that is to $\partial D/\partial \boldsymbol{n}$. It is difficult to visualise a complex ray, normal to a complex refractive index surface, in the six-dimensional space of the real and imaginary parts of $\boldsymbol{n}$, and we have to be content with these formulae. On this basis the whole of the theory of the canonical equations of a ray, §§ 14.2–14.7, can be carried through for complex $\boldsymbol{n}$, x, y, z.

The first method used in § 5.3 cannot, however, immediately be taken over for media with losses. It used the time averaged Poynting vector $\mathbf{\Pi}_{av}$ which is necessarily a real vector in real space and cannot have a complex direction. We need to find a complex analogue of $\mathbf{\Pi}_{av}$ that has the direction of the complex ray. It will now be shown that the required vector is the bilinear concomitant vector $\boldsymbol{W}$. Its z component W in a stratified medium has already been used in §7.14, and defined by (7.89). The name comes from its use in the theory of differential equations. The following treatment is for monochromatic waves, that is for 'steady state' conditions.

Consider a fictitious adjoint medium identical with the original medium except that at each point the direction of the superimposed magnetic field is reversed. For an electron plasma this means that the direction cosines of $\boldsymbol{Y}$ are reversed in sign, so that any term containing an odd power of Y as a factor is reversed in sign. For the more general plasma it means that ε_1 and ε_2 are interchanged. Thus for an electron plasma (3.35) shows that the susceptibility matrix $\boldsymbol{M}$ and thence the dielectric constant $\boldsymbol{\varepsilon} = 1 + \boldsymbol{M}$ is transposed. Similarly, for the more general plasma, (3.53) or (3.54) or (3.55) shows that $\boldsymbol{\varepsilon}$ is transposed. Each of these forms is for a particular set of axes but the property still holds when $\boldsymbol{\varepsilon}$ is transformed to use any other set of Cartesian axes. Hence, in the subscript notation of (2.54), (5.27) its adjoint form is given by

$$\bar{\varepsilon}_{ij} = \varepsilon_{ji}. \tag{14.97}$$

In the adjoint medium there is a source of waves identical with the actual source in the original medium. The wave fields in the adjoint medium are the adjoint fields $\bar{\boldsymbol{E}}$, $\bar{\boldsymbol{\mathscr{H}}}$. An example for a stratified medium was given in §7.14 item (4). They must satisfy the differential equations that are the adjoint of Maxwell's equations (2.44) for monochromatic waves. These are formed by reversing the sign of k and thence of $\omega = kc$; compare (7.95). In the constitutive relation, however, the sign of ω is not reversed. The dispersion relation at any point is the same in the original and the adjoint medium. Thus the refractive index surfaces and ray directions and ray paths are the same in the two media. For any point Q on a ray in the original medium, each component of the predominant field contains a factor

$$\exp\{\mathrm{i}(\omega t - kP)\} = \exp\left(\mathrm{i}\omega t - \mathrm{i}k\int \mathscr{M}\,\mathrm{d}s\right) \tag{14.98}$$

where P is the phase path, and $\mathscr{M} = n\cos\alpha$ is the ray refractive index; see (5.35), (14.42). Here ds is an element of the ray path, and the integral is evaluated along the ray up to the point Q. In a medium with losses, $\mathrm{Im}(\mathscr{M})$ and $\mathrm{Im}(P)$ are negative and in real space the wave is attenuated as it travels. For the corresponding fictitious ray in the adjoint medium the factor is $\exp\{\mathrm{i}(-\omega t + kP)\}$. Now P is the same in the two cases so that the wave is enhanced as it travels. This wave has no physical existence and is merely a mathematical adjunct. For a loss-free medium and real space, P is real, and the waves are neither attenuated nor enhanced, respectively.

Now, by analogy with (7.97), the bilinear concomitant vector is defined as follows:

$$\boldsymbol{W} = \boldsymbol{E} \wedge \bar{\mathscr{H}} + \bar{\boldsymbol{E}} \wedge \mathscr{H}. \tag{14.99}$$

Its use for radio propagation problems was suggested by Budden and Jull (1964). A full theory was given by Suchy and Altman (1975a).

In the theory of the dispersion relation for a loss-free medium, and in real space, i appears only where it multiplies Y, for example in (3.35) or (4.12). For the more general plasma i appears only where it multiplies $(\varepsilon_1 - \varepsilon_2)$, for example in (3.53) or (4.61). Thus reversal of the direction of the superimposed magnetic field has the same effect as reversing the sign of i. In this case, therefore, the adjoint fields $\bar{\boldsymbol{E}}, \bar{\mathscr{H}}$ are proportional to the complex conjugates $\boldsymbol{E}^*$, $\mathscr{H}^*$, and $\boldsymbol{W}$ is proportional to $\boldsymbol{\Pi}_{\text{av}}$ (2.63). But $\boldsymbol{E}^*$, $\mathscr{H}^*$ are not analytic functions of x, y, z, n and so the process of analytic continuation cannot be applied to them. On the other hand $\bar{\boldsymbol{E}}, \bar{\mathscr{H}}$ are analytic functions. They and $\boldsymbol{W}$ can be continued analytically into complex space, and used for media with losses.

The adjoint fields $\bar{\boldsymbol{E}}, \bar{\mathscr{H}}$ have been defined as the fields in the adjoint medium from a source identical to the actual source, but this is not always the best choice. They must satisfy the adjoint of Maxwell's equations which are linear and homogeneous, so that $\bar{\boldsymbol{E}}, \bar{\mathscr{H}}$ can be multiplied by any constant. The choice of this constant is a matter of convenience. This problem has been discussed for a stratified medium in §7.14(1), (7), (8), where the constant was chosen so that, for a real ray in a loss-free part of the medium, $\bar{\boldsymbol{E}}$, $\bar{\mathscr{H}}$ are the same as $\boldsymbol{E}^*$, $\mathscr{H}^*$. The same choice is permissible for a medium that is not plane stratified. But if this applies in one loss-free region, it does not apply for the same ray in a second loss-free region separated from the first by a region with losses, or by a surface where the refractive index shows a resonance. This is important in the theory of resonance tunnelling (Budden, 1979); see §19.6. The choice of the constant is the problem of the normalisation of $\bar{\boldsymbol{E}}, \bar{\mathscr{H}}$. It does not affect the theory in the rest of this section.

Now it can be shown that $\boldsymbol{W}$ is normal to the refractive index surface. The proof follows exactly the steps from (5.20) to (5.28) but with the following changes:

(a) $\boldsymbol{n}$ is not restricted to be real;

(b) the adjoints $\bar{\boldsymbol{E}}, \bar{\mathscr{H}}$ are used instead of the complex conjugates $\boldsymbol{E}^*, \mathscr{H}^*$;

(c) the adjoints of (5.20), (5.21) are used instead of their complex conjugates, for substituting in (5.24), (5.25);

(d) $\boldsymbol{\Pi}_{\text{av}}$ is replaced by $\boldsymbol{W}$.

(e) (14.97) is used in the stage (5.27).

Hence $\boldsymbol{W}$ is in the same direction as the ray. This applies for complex rays and in complex space.

These results have been established for monochromatic waves. Thus the only time dependence is through the factor $\exp(i\omega t)$ for $\boldsymbol{E}$, $\mathscr{H}$, and $\exp(-i\omega t)$ for their adjoints, and (14.99) shows that $\boldsymbol{W}$ is independent of time. From (14.99) together with Maxwell's equations (2.44) and their adjoints it can be shown that

$$\operatorname{div} \boldsymbol{W} = 0 \tag{14.100}$$

so that $\boldsymbol{W}$ is solenoidal and must be a flux of some indestructible quantity that exists in the medium with a density $\mathscr{Q}$ per unit volume. For a loss-free medium $\boldsymbol{W}$ is $4Z_0\boldsymbol{\Pi}_{\mathrm{av}}$ so that, when steady state conditions do not hold, $\operatorname{div} \boldsymbol{W} = -4Z_0\partial\mathscr{S}/\partial t$ where $\mathscr{S}$ is the total energy per unit volume; § 2.12. In a medium with losses $\mathscr{Q}$ is a complex generalisation of $4Z_0\mathscr{S}$.

The energy flux in a ray pencil in real space and in a loss-free medium was studied in § 10.17 and it was shown that the signal intensity is given by (10.47). This method can now be applied to a medium with losses and to use the flux of $\mathscr{Q}$ in a ray pencil of complex rays. The same expression (10.47) is obtained but it is now complex, so that it cannot represent the signal intensity, which is real. It can be shown, however, that (10.47) can still be used if the ray pencil enters a region of real space where the medium is loss-free. Then the signal intensity is found by taking the modulus of (10.47) and multiplying by $\exp\{2\,\mathrm{Im}(P)\}$ where P is the phase path. The proof is given by Budden and Terry (1971, Appendix A).

In the foregoing discussion an attempt has been made to explain the main properties of complex rays and the bilinear concomitant vector in anisotropic absorbing media. The theory has been carried considerably further in some recent papers. For example Suchy (1982) has studied the vector $\boldsymbol{W}$ where steady state conditions do not hold. Then $\mathscr{Q}$ varies with time t and (14.100) must be replaced by

$$\operatorname{div} \boldsymbol{W} = -\partial\mathscr{Q}/\partial t. \tag{14.101}$$

The theory has also been applied to media with spatial dispersion (Suchy, 1982).

14.13. Reciprocity and nonreciprocity with rays 1. The aerial systems

The reciprocity theorem of electromagnetism is not in general true for systems that include a gyrotropic medium, and so it does not apply to radio signals propagated through the ionosphere or magnetosphere. But in certain special cases there is reciprocity for a radio-communication link. The theory for the general case is discussed in § 18.12. In this and the two following sections it is examined when the approximations of ray theory are used, as in this chapter. The discussion is based on the more detailed treatment of Budden and Jull (1964).

When a signal travels along a ray, the path of the ray is unaltered if the direction of the wave normal is everywhere reversed. This was proved in § 10.5 for a stratified medium. In the more general case it follows from the properties of the refractive index surface, because the direction of the ray is everywhere perpendicular to the local refractive index surface. This has a centre of symmetry at the origin so that if the wave normal is reversed, the ray is then simply antiparallel to its original direction. The attenuation and phase change of a signal travelling along a ray are given by $\exp(-ik\int \mathscr{M}\, ds)$ where ds is an element of the ray path and $\mathscr{M} = \boldsymbol{n}\cos\alpha$ is the ray refractive index (14.42). This still applies when collisions are allowed for so that $\mathscr{M}$ is complex, and it applies for complex rays. But $\mathscr{M}$ is unchanged by a reversal of the ray direction. Thus the ray path, and the attenuation and phase change, are the same for a signal travelling either way along the path. There is an exception to this. It has been assumed that the rays can be described by the eikonal function (14.7), that is the 'phase memory' term, and that this depends on the refractive index vector $\boldsymbol{n}$. It is this term that has been used to find the ray direction, and the ray refractive index $\mathscr{M}$. Cases can occur where this is not accurate enough and another factor must be included in the eikonal. It is called 'additional memory' and is explained in § 16.9. It has the effect that the ray path from A to B is not the same as the path from B to A. It is a small effect especially at high frequencies, and in this chapter it is ignored.

We consider two aerials at fixed points A, B in free space though not necessarily on the ground; one or both could be on space vehicles. Each aerial is a linear electrical system that does not contain nonreciprocal circuit elements. When one aerial is used as a transmitter, a pencil of rays diverges from it. A small number, usually one or two, of these reaches the other aerial, and these are called the 'principal rays'. The divergence of the neighbouring rays affects the amplitude of the received signal, as discussed in §§ 10.17, 14.8. If the other aerial is now used as the transmitter, the principal rays follow the same paths as before, in the opposite direction, but their neighbouring rays follow different paths. The effect on the received signal amplitude of the divergence of the rays in a ray pencil is given, for a stratified medium, by (10.43)–(10.47). These contain integrals (10.43) along the ray path and it is easy to show that they are unchanged if the direction of the ray is reversed. For the more general case the signal amplitude is found by integrating the Buckley equations (14.71), (14.72) along the ray path, and again the result is unchanged if the direction is reversed.

When nonreciprocity occurs, therefore, it has nothing to do with the ray path, nor with the attenuation and phase shift, nor with the divergence of the rays in the pencils. It arises because of the interaction of the waves with the transmitting and receiving aerials, and depends on the polarisations of the waves in the signal, and of the waves that the aerials can radiate.

At any point on a ray within the plasma, the ray direction, the wave normal and the earth's magnetic field are coplanar, §§ 5.3, 5.4, and the plane containing them will be called the 'magnetic plane'. At each point on the ray, therefore, we define a right-handed Cartesian axis system such that any vector, for example the electric intensity $\boldsymbol{E}$, has components E_1, E_2, E_3 where E_3 has the direction of the wave normal for a signal going from A to B, and E_1 is in the magnetic plane. Thus the 1, 2, 3 axes are the same as the x, y, z axes of ch. 4, but not of later chapters. For a signal going from B to A the *same* axes are used so that E_3 is now *opposite* to the wave normal.

In the approximations of ray theory, the fields at a point on the ray are the same as those for a progressive plane wave, ordinary or extraordinary, in a homogeneous medium. For this wave, as in ch. 4, we use n for the refractive index, ρ for the polarisation E_2/E_1, and L for the ratio E_3/E_1 as given by (4.54). Subscripts O, E or superscripts (O), (E) will be attached later when we need to distinguish the ordinary and extraordinary waves. Then for a signal going from A to B, the fields in this axis system are

$$\boldsymbol{E} = (1, \rho, L)E_1, \quad \mathscr{H} = (-n\rho, n, 0)E_1. \tag{14.102}$$

For a ray going from B to A, the sign of n is reversed, but ρ and L are unchanged; see end of § 4.4.

For points outside the plasma the ray is straight. The ray and the wave normal coincide and give the direction of the 3-axis. When a ray emerges into free space, n_{O} and n_{E} become nearly equal since both are near to unity, and there is a cumulative coupling process that causes failure of the ray theory approximations. This is the process of 'limiting polarisation', §§ 17.10, 17.11. Suppose that the emerging wave is an ordinary wave within the plasma. After emerging it is no longer purely ordinary but attains a constant limiting polarisation $\check{\rho}_{\mathrm{O}}$. Similarly an extraordinary wave attains a limiting polarisation $\check{\rho}_{\mathrm{E}}$. For the ray through A these values are very close to the values ρ_{O}, ρ_{E} for a point P on the ray just within the plasma. The exact position of P is unimportant because, when X is very small, ρ_{O} and ρ_{E} are insensitive to small changes of X. Hence to a good approximation we expect that

$$\check{\rho}_{\mathrm{O}}\check{\rho}_{\mathrm{E}} \approx 1 \tag{14.103}$$

from (4.29). On the part AP of the ray in free space, the direction of the 1 axis is parallel to the magnetic plane at P. This ensures that (14.103) is true at A. A similar argument applies for the ray through B.

In discussions of reciprocity we consider two experiments. In the first a voltage V is applied to the terminals of the aerial at A and the current i_1 at the short-circuited terminals of B is measured. In the second the same voltage V is applied to the terminals of the aerial at B and the current i_2 at the short-circuited terminals of the aerial at A is measured. If the electromagnetic fields are entirely within isotropic media, then $i_1 = i_2$, and this is the standard reciprocity theorem. It is discussed further in § 18.12 (see also, for example, Landau and Lifshitz, 1960; Ginzburg, 1970; Collin and Zucker, 1969). It is generally true for full wave solutions and is not restricted to ray theory. When anisotropic media are present the theorem is not generally true, but it may still be true in some special cases. There are other special cases where $i_1 = -i_2$ and this will be called 'antireciprocity'. Examples are given in §§14.15, 18.12.

To apply these ideas we next define certain constants of the aerials when they are used for conveying a signal along a particular path AB in either direction. First let the aerial at A be used as the transmitter, by applying to its terminals a voltage V at the signal frequency. Let E_1, E_2 be the resulting components of $\boldsymbol{E}$ at a point on the ray AB at distance D from A. This point must be in free space, but D must be large enough to ensure that the storage fields of the aerial are negligible compared with the radiation field. Then

$$E_1 = \gamma_A V \cos\psi_A, \quad E_2 = \gamma_A V \sin\psi_A \tag{14.104}$$

where γ_A, ψ_A are constants of the aerial. If it radiates linearly polarised waves, ψ_A is real and gives the plane of $\boldsymbol{E}$, and we shall then say that the aerial is linearly polarised in this plane. But in general the aerial radiates an elliptically polarised wave and ψ_A is complex. Similarly, if the aerial at B is used as the transmitter by applying the voltage V to its terminals, then the fields on the ray BA at distance D from B in free space are

$$E_1 = \gamma_B V \cos\psi_B, \quad E_2 = \gamma_B V \sin\psi_B. \tag{14.105}$$

If the aerial at A is used as a receiver, and the signal reaching it along the ray BA has electric field components E_1', E_2' at A, then the current at the short-circuited terminals at A is

$$i_2 = K\gamma_A(E_1' \cos\psi_A + E_2' \sin\psi_A). \tag{14.106}$$

If the aerial at B is used as a receiver, and if the electric field components there are E_1'', E_2'' then the current at the short-circuited terminals at B is

$$i_1 = K\gamma_B(E_1'' \cos\psi_B + E_2'' \sin\psi_B). \tag{14.107}$$

The constant K is a property of the aerials only, and not of the fields. It can be proved that the same K must be used in the two cases, by supposing that the path AB is entirely in free space and using the standard reciprocity theorem.

14.14. Reciprocity and nonreciprocity with rays 2. The electric and magnetic fields

Consider now one magnetoionic component, which will be taken to be the ordinary wave. The whole of the following discussion applies if the roles of the ordinary and extraordinary waves are reversed. For a point on a ray going from A to B, the field components are, from (14.102)

$$\boldsymbol{E} = (1, \rho_O, L_O)E_1^{(O)} \quad \mathscr{H}(-n_O\rho_O, n_O, 0)E_1^{(O)}. \tag{14.108}$$

The components of the corresponding adjoint field were defined in §14.12. They have the same ratios as the fields in a fictitious medium in which the direction of the earth's magnetic field is reversed, and the wave normal and ray directions are unchanged. Hence it can be shown from (4.54) and (4.20) that n_O, L_O are unchanged, and the sign of ρ_O is reversed (see problem 7.4). Then for the adjoint fields

$$\bar{\boldsymbol{E}} = (1, -\rho_O, L_O)\bar{E}_1^{(O)}, \quad \bar{\mathscr{H}} = (n_O\rho_O, n_O, 0)\bar{E}_1^{(O)}. \tag{14.109}$$

The bilinear concomitant vector (14.99) can now be found and is

$$\boldsymbol{W} = -2n_O(L_O, 0, \rho_O^2 - 1)E_1^{(O)}\bar{E}_1^{(O)}. \tag{14.110}$$

Its (complex) length is

$$W = -2n_O\{L_O^2 + (\rho_O^2 - 1)^2\}^{\frac{1}{2}}E_1^{(O)}\bar{E}_1^{(O)}. \tag{14.111}$$

It was shown in § 14.12 that $\boldsymbol{W}$ everywhere has the same direction, in general complex, as the ray.

Now (14.100) shows that div $\boldsymbol{W} = 0$, so that the flux of $\boldsymbol{W}$ across any closed surface is zero. In particular, if the surface is a very narrow tube whose sides are everywhere parallel to rays emanating from A, there is no flux of $\boldsymbol{W}$ across the sides so that the total flux for any cross section must be the same at all points of the tube. Such a tube will be called a 'ray tube'. Let a denote the area of any cross section normal to the axis of the tube. Then Wa is constant along a ray, whence it follows from (14.111) that

$$n_O a\{L_O^2 + (\rho_O^2 - 1)^2\}^{\frac{1}{2}}E_1^{(O)}\bar{E}_1^{(O)} = \text{constant} \tag{14.112}$$

where an unimportant factor -2 is omitted. Now, for any point on the ray, $E_1^{(O)}$ contains a factor $\exp\{\mathrm{i}(\omega t - kP_A)\}$, from (14.98), where P_A is the phase path from A to the point considered. Similarly $\bar{E}_1^{(O)}$ contains a factor $\exp\{\mathrm{i}(-\omega t + kP_A)\}$. These factors cancel in (14.112). Let

$$F_O = E_1^{(O)}n_O^{\frac{1}{2}}\{L_O^2 + (\rho_O^2 - 1)^2\}^{\frac{1}{4}} \tag{14.113}$$

and similarly with $\bar{F}_O$, $\bar{E}_1^{(O)}$ for the adjoints. Then $F_O\bar{F}_O$ is a complex generalisation of the signal intensity Π used in § 14.8. Now on the ray

$$F_O a^{\frac{1}{2}} = \text{constant} \times \exp(-\mathrm{i}kP_A) \tag{14.114}$$

where the factor $\mathrm{e}^{\mathrm{i}\omega t}$ is omitted as usual.

Let a voltage V be applied to the terminals of the transmitting aerial at A. The fields E_1, E_2 at a distance D away in free space, on the ray, are given by (14.104) and must be resolved into components that will become ordinary and extraordinary after they enter the plasma. Then

$$E_1 = E_1^{(O)} + E_1^{(E)}, \quad E_2 = E_2^{(O)} + E_2^{(E)} = \rho_{OA}E_1^{(O)} + E_1{}^{(E)}/\rho_{OA} \tag{14.115}$$

where ρ_{OA} is the limiting value $\breve{\rho}_O$ of ρ_O on the ray through A, and (14.103) has been used. In the free space $L_O = 0$, $n_O = 1$. Hence at the same point, from (14.113) and (14.104)

$$\begin{aligned} F_O &= (\rho_{OA}E_2 - E_1)(\rho_{OA}^2 - 1)^{-\frac{1}{2}} \\ &= \gamma_A V(\rho_{OA}\sin\psi_A - \cos\psi_A)(\rho_{OA}^2 - 1)^{-\frac{1}{2}} \end{aligned} \tag{14.116}$$

which is the value at a distance D from A. The signal now travels along the ray AB through the anisotropic plasma and reaches B. To find F_O at B we must multiply by $J_O \exp\{-ik(P_{AB}-D)\}$ where P_{AB} is the phase path from A to B. J_O is determined solely by the geometry of the ray system and allows for the divergence of the rays from A. Hence at B, F_O is

$$F_{OB} = J_O \gamma_A V(\rho_{OA} \sin\psi_A - \cos\psi_A)(\rho_{OA}^2 - 1)^{-\frac{1}{2}} \exp\{-ik(P_{AB}-D)\} \tag{14.117}$$

and from (14.113) the fields E_1, E_2 at B are

$$E_1'' = F_{OB}(\rho_{OB}^2 - 1)^{-\frac{1}{2}}, \quad E_2'' = \rho_{OB} F_{OB}(\rho_{OB}^2 - 1)^{-\frac{1}{2}}. \tag{14.118}$$

These are now inserted in (14.107) to give the current i_1 at the short-circuited terminals of the aerial at B:

$$i_1 = M_O(\rho_{OA} \sin\psi_A - \cos\psi_A)(\cos\psi_B + \rho_{OB} \sin\psi_B) \tag{14.119}$$

where

$$M_O = J_O K \gamma_A \gamma_B V(\rho_{OA}^2 - 1)^{-\frac{1}{2}}(\rho_{OB}^2 - 1)^{-\frac{1}{2}} \exp\{-ik(P_{AB}-D)\}. \tag{14.120}$$

Next let the aerial at B be used as the transmitter. The divergence factor J_O in (14.117) is the same as before for the reasons given in § 14.13, and the exponential factor is also the same. The only difference is that the subscripts A and B are interchanged, so M_O is unchanged. The current at the short-circuited terminals of the aerial at A is

$$i_2 = M_O(\cos\psi_A + \rho_{OA} \sin\psi_A)(\rho_{OB} \sin\psi_B - \cos\psi_B). \tag{14.121}$$

14.15. Reciprocity and nonreciprocity with rays 3. Applications

Suppose that there is a communication link between A and B with only one ordinary ray. Then the condition $i_1 = i_2$ for reciprocity gives, from (14.119), (14.121):

$$\rho_{OA} \sin\psi_A \cos\psi_B = \rho_{OB} \cos\psi_A \sin\psi_B. \tag{14.122}$$

Similarly the condition $i_1 = -i_2$ for antireciprocity gives:

$$\cos\psi_A \cos\psi_B = \rho_{OA}\rho_{OB} \sin\psi_A \sin\psi_B. \tag{14.123}$$

In many practical cases the limiting polarisations are very nearly circular. This is because ρ_{OA}, ρ_{OB} are insensitive to the angle Θ between the earth's magnetic field and the wave normal, except when it is close to $\pm\frac{1}{2}\pi$. If this transverse case is excluded, it is found that ρ_{OA}, ρ_{OB} are both close to $\pm$i. For ionospheric ray paths between ground stations that are fairly close together, Θ is acute for the ray near A and obtuse for the ray near B or vice versa. Then ρ_{OA}, ρ_{OB} are equal with opposite signs.

If $\rho_{OA} = -\rho_{OB}$, the reciprocity condition (14.122) is

$$\psi_A + \psi_B = 0. \tag{14.124}$$

Thus there is reciprocity when both aerials are linearly polarised in the magnetic plane, and the condition still holds if the aerials are now rotated through equal angles but in opposite senses about the local direction of the ray AB.

If ρ_{OA} and ρ_{OB} satisfy the further condition $\rho_{OA} = -\rho_{OB} = \pm i$, the antireciprocity condition (14.123) is

$$\psi_A - \psi_B = \tfrac{1}{2}\pi. \tag{14.125}$$

Thus there is antireciprocity when the aerials are linearly polarised, one in the magnetic plane and the other at right angles to it. The condition still holds if the aerials are rotated from these positions through equal angles and in the same sense about the local direction of the ray AB. These conclusions all apply equally well if the one ray is an extraordinary ray. For further discussion see Budden and Jull (1964).

It often happens that a signal goes over two different ray paths simultaneously, and the commonest case is when one ray is ordinary and the other extraordinary. For the second ray the subscript O in (14.119)–(14.121) must be replaced by E. The two rays may leave the aerials in slightly different directions but in practical cases we expect $\gamma_A \psi_A \gamma_B \psi_B$ to be the same for both rays. Then the contributions to i_1, i_2 from the two signals are

$$\left.\begin{aligned} i_1 &= M_O(\rho_{OA}\sin\psi_A - \cos\psi_A)(\cos\psi_B + \rho_{OB}\sin\psi_B) \\ &\quad + M_E(\rho_{EA}\sin\psi_A - \cos\psi_A)(\cos\psi_B + \rho_{EB}\sin\psi_B), \\ i_2 &= M_O(\cos\psi_A + \rho_{OA}\sin\psi_A)(\rho_{OB}\sin\psi_B - \cos\psi_B) \\ &\quad + M_E(\cos\psi_A + \rho_{EA}\sin\psi_A)(\rho_{EB}\sin\psi_B - \cos\psi_B). \end{aligned}\right\} \tag{14.126}$$

In rarer cases there could be three or more ray paths (see, for example, § 12.7 and fig. 12.9) and further terms could be added to deal with this.

These results display the following properties, which apply for any number of rays.

(a) There is reciprocity when $\psi_A = \psi_B = 0$, that is when both aerials are linearly polarised in the magnetic plane.

(b) There is reciprocity when $\psi_A = \psi_B = \frac{1}{2}\pi$, that is when both aerials are linearly polarised at right angles to the magnetic plane.

(c) There is antireciprocity when $\psi_A = 0$, $\psi_B = \pm\frac{1}{2}\pi$, or when $\psi_B = 0$, $\psi_A = \pm\frac{1}{2}\pi$, that is when both aerials are linearly polarised, one in the magnetic plane and the other at right angles to it.

These three examples are the results of a more general treatment of reciprocity and can be proved without recourse to ray theory; see §18.12.

An important case that often occurs in practice is for medium and high frequencies when there are two rays from A to B, one ordinary and the other extraordinary. The signal is assumed to be continuous, not pulsed or otherwise modulated. The two waves are of roughly equal amplitudes which means that the attenuations are small and about the same for the two paths. The phase paths, however, are different and change with time because of slow movements or changes of the ionosphere. Hence the difference of the phase paths changes with time so that the two contributions to the received signal have a variable phase difference, and alternately reinforce each other or interfere destructively. This gives a very regular variation of the signal called 'polarisation fading' because its occurrence depends on the different polarisations of the ordinary and extraordinary waves. To study it, assume that

$$\rho_{OA} = -\rho_{OB} = \mathrm{i}, \quad \rho_{EA} = -\rho_{EB} = -\mathrm{i} \tag{14.127}$$

which is typical for propagation between two ground stations at temperate latitudes in the same hemisphere. The divergence factors J_O, J_E are different but it would be expected that they are roughly equal in most practical cases. Then M_O, M_E are the same except for the exponential factor. Let

$$M_O = M_E \exp\{2\mathrm{i}\phi(t)\} \tag{14.128}$$

where $2\phi(t)/k$ is the difference between the two phase paths. Then (14.126) gives

$$\begin{aligned} i_1 &= -M_E[\exp\{\mathrm{i}(2\phi - \psi_A - \psi_B)\} + \exp\{\mathrm{i}(\psi_A + \psi_B)\}], \\ i_2 &= -M_E[\exp\{\mathrm{i}(2\phi + \psi_A + \psi_B)\} + \exp\{-\mathrm{i}(\psi_A + \psi_B)\}]. \end{aligned} \tag{14.129}$$

Let both aerials be linearly polarised so that ψ_A and ψ_B are real. Then

$$|i_1| = M|\cos(\psi_A + \psi_B - \phi)|, \quad |i_2| = M|\cos(\psi_A + \psi_B + \phi)| \tag{14.130}$$

where M is a real constant. The cosine factors give the polarisation fading. Now let

$$\psi_A + \psi_B = \tfrac{1}{2}r\pi \tag{14.131}$$

where r is any integer or zero. Then the fading for the direction A → B is exactly in step with that for B → A. This occurs for the two cases r even which gives reciprocity, cases (a) and (b) above, and r odd which gives antireciprocity, case (c). Alternatively let

$$\psi_A + \psi_B = (\tfrac{1}{2}r \pm \tfrac{1}{4})\pi. \tag{14.132}$$

Now the cosine factors in (14.130) are exactly out of step since one is a maximum when the other is zero. The result of using other values of ψ_A and ψ_B can be found from (14.130) in a similar way. These and other effects have been confirmed by observations; see Jull and Pettersen (1962, 1964).

These results have an alternative explanation. The two rays reaching a receiver have been assumed to be roughly circularly polarised with opposite senses and equal amplitudes. They therefore combine to give a linearly polarised wave whose plane undergoes Faraday rotation when the phase difference changes; see §13.7. The direction of this rotation is in opposite senses about the ray direction AB for signals going in the two directions. The fading occurs as the angle between the plane of polarisation of the wave and of the aerial varies. This angle can be expressed in terms of $\psi_A, \psi_B, \phi(t)$ and the above results then follow. See Budden and Jull (1964) for further details.

PROBLEMS 14

14.1. Show by a counter example that, for different rays through the same point, the directions of the effective stratification vector (defined near end of § 14.3) are not necessarily the same.

Hint: A possible counter example is as follows. Consider a collisionless electron plasma in which the electron concentration is independent of the coordinates, but the magnetic field varies in space so that $Y_x = Y_0 2^{-\frac{1}{2}} + \alpha z$, $Y_z = Y_0 2^{-\frac{1}{2}} + \alpha x, Y_y = 0$. The dispersion relation (4.65), (4.66) can be written

$$D \equiv (1 - X - Y^2)n^4 - 2(1-X)^2 n^2 + (1-X)^3 + Y^2\{(2-X)n^2 - 1 + X\} \\ + X(n^2 - 1)(\boldsymbol{Y}\cdot\boldsymbol{n})^2 = 0.$$

Consider two ordinary rays through the origin with wave normals parallel (a) to the x-axis, (b) to the z-axis. The angles between $\boldsymbol{Y}$ and the wave normals are the same so the two refractive indices n are the same. The two ray directions are in the n_x–n_z plane and make equal and opposite angles with $\boldsymbol{Y}$. Show that the angles between the effective stratification vector and the x axis are (a) $\arctan(n^2-1)(1-X)/(n^2-1+X)$, (b) $\arctan(n^2-1+X)/\{(n^2-1)(1-X)\}$ and so not in general the same.

14.2. It is required to find a ray that leaves a transmitter at the origin and goes to a receiver at the point $(x_0, 0, 0)$. For a particular ray at the transmitter the polar angles θ, ϕ of the wave normal are θ_i, ϕ_i and this ray goes through the point $(x_i, y_i, 0)$. This applies for three different rays, $i = 1$ to 3. By linear interpolation, that is by a two-dimensional form of the rule of false position (see (14.1)), show that estimates of the required values θ_0, ϕ_0, are

$$\theta_0 \approx \begin{vmatrix} \theta_1 & x_1 & y_1 & 1 \\ \theta_2 & x_2 & y_2 & 1 \\ \theta_3 & x_3 & y_3 & 1 \\ 0 & x_0 & 0 & 1 \end{vmatrix} \times \begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & x_3 & 1 \end{vmatrix}^{-1}$$

and a similar expression with ϕs instead of θs.

14.3. The dispersion relation may be written $\omega(x, y, z; \kappa_x, \kappa_y, \kappa_z) = \Omega$ where $\boldsymbol{\kappa} = \omega\boldsymbol{n}/c$ and Ω is the angular frequency of the wave. The first three canonical ray tracing equations are $\mathrm{d}\boldsymbol{r}/\mathrm{d}\tau = A\partial\omega/\partial\boldsymbol{n}$. Let $\omega A\,\mathrm{d}\tau/c = \mathrm{d}t$. Use these equations to prove that the group velocity $\mathcal{U} = \partial\omega/\partial\boldsymbol{\kappa} = \mathrm{d}\boldsymbol{r}/\mathrm{d}t$ and show that $1/A = \boldsymbol{n}\cdot\mathcal{U}\omega/c$.

14.4. For a high frequency communication link between two points A and B there is only one ray, that goes through the ionosphere. Its limiting polarisations at A and B are ρ_A, ρ_B where $\rho_A = -\rho_B = \mathrm{i}$. The aerials at A and B are both designed to emit only this wave. Show that the signal emitted from A cannot be received at B, nor that emitted from B at A. The aerial at B is now changed so that it can receive the

signal from A. Show that now a signal from A can be received at B, but a signal from B cannot be received at A. Discuss how to design the aerials so that communication in both directions is possible, whatever may be the values of ρ_A, ρ_B.
(Note: to attain the required condition $E_2/E_1 = \mathrm{i}$ at A in (14.104) it is necessary that $\psi_A \to \mathrm{i}\infty$, but $\gamma_A \cos\psi_A$, $\gamma_A \sin\psi_A$ tend to bounded limits).