8

The Airy integral function and the Stokes phenomenon

8.1. Introduction

In ch. 7 it was shown that at most levels in a slowly varying stratified ionosphere, and for radio waves of frequency greater than about 100 kHz, the propagation can be described by approximate solutions of the differential equations, known as W.K.B. solutions. The approximations fail, however, near levels where two roots q of the Booker quartic equation approach equality. In this chapter we begin the study of how to solve the differential equations when this failure occurs. For an isotropic ionosphere there are only two values of q and they are equal and opposite, and given by (7.1), (7.2). The present chapter examines this case. It leads on to a detailed study of the Airy integral function, which is needed also for the solution of other problems. Its use for studying propagation in an anisotropic ionosphere where two qs approach equality is described in § 16.3.

For an isotropic ionosphere, a level where $q = 0$ is a level of reflection. In § 7.19 it was implied that the W.K.B. solution for an upgoing wave is somehow converted, at the reflection level, into the W.K.B. solution for a downgoing wave with the same amplitude factor, and this led to the expression (7.151) for the reflection coefficient R. The justification for this assertion is examined in this chapter and it is shown in § 8.20 to require only a small modification, as in (7.152).

8.2. Linear height distribution of electron concentration and isolated zero of q

The coordinate system used here is as defined in § 6.1.

If electron collisions are neglected, q^2 is related to the electron concentration N thus

$$q^2 = C^2 - X, \tag{8.1}$$

where X is given by (3.5) and is directly proportional to N, and $C = \cos\theta$ where θ is the angle between the incident wave normal and the vertical. Thus q depends on z through X, and is zero when $X = C^2$.

The simplest example of a zero of q occurs when X is a slowly increasing monotonic function of z, as shown in fig. 8.1(a). Then q^2 is a decreasing monotonic function of z as shown in fig. 8.1(b). It is zero where $z = z_0$, and to a first approximation the variation of q^2 with z may be taken as linear near z_0, so that

$$q^2 \approx -a(z - z_0), \tag{8.2}$$

where a is a constant. The right side of (8.2) may be regarded as the first term in the Taylor expansion for q^2 about the point z_0. It was shown in § 7.10 that the W.K.B. solutions in this case are good approximations provided that $|z - z_0|$ exceeds a certain minimum value, M, say. In the present chapter up to § 8.19 it is assumed that (8.2) may be used for q^2 for all values of $|z - z_0|$ from zero up to M. Within this range the solution of the differential equations can be expressed in terms of Airy integral functions, and outside the range the W.K.B. solutions are so chosen that they fit continuously to the solution within the range.

If q^2 is given exactly by (8.2) there is only one value of z, namely z_0 which makes

Fig. 8.1. Dependence on the height z of (a) X (proportional to electron concentration) and (b) q^2, for a slowly varying ionosphere when q^2 has an 'isolated' zero.

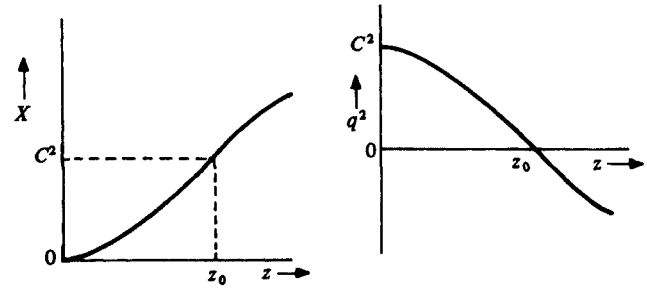
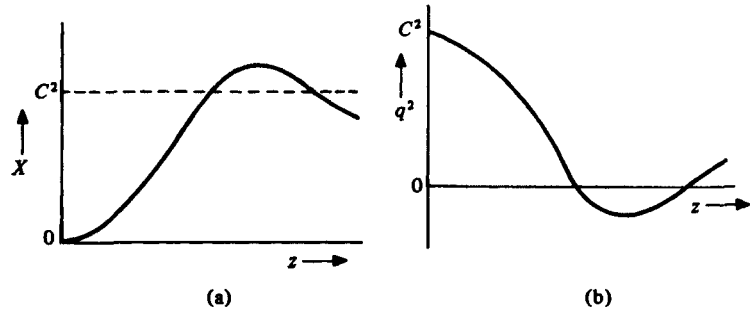


Fig. 8.2. Dependence on the height z of (a) X (proportional to electron concentration) and (b) q^2, when q^2 has two zeros close together.



$q = 0$. If, however, the line in fig. 8.1(b) is slightly curved where $z = z_0$ then, in a range of z near z_0, (8.2) may be replaced by

$$q^2 \approx -a(z - z_0) + b(z - z_0)^2, \tag{8.3}$$

where b determines the curvature at $z = z_0$. Now q is zero both at $z = z_0$ and at $z = z_0 + a/b$. This case is illustrated in fig. 8.2(a, b), and could occur, for a frequency less than the penetration frequency, when the electron concentration has a maximum value. If b is very small in (8.3), the second zero is at a great distance a/b from the first, provided that a is not small.

Hence the condition that the variation of q^2 with z shall be nearly linear, near a zero of z, is equivalent to saying that this zero is at a great distance from the next nearest zero. In other words, the zero of q must be isolated. A more exact statement of this condition is given by (8.75), § 8.20.

If a is very small, the two zeros of (8.3) can be close together even when b is small. This case could occur when the frequency is only slightly less than the penetration frequency. Here the W.K.B. solutions fail for a range of z which includes both zeros. It is then necessary to study the differential equations when q^2 varies according to the parabolic law. This leads to the phenomena of partial penetration and reflection, and is discussed in § 15.10. For frequencies slightly greater than the penetration frequency, q has no zeros at any real value of z, but there are zeros in the complex z plane. This case also is discussed in §§ 15.9, 15.10.

Another form of variation of q with z that will be encountered (§ 19.6) is

$$q^2 = \alpha \frac{z - z_0}{z - z_1}, \tag{8.4}$$

where α, z_0, z_1 are constants. Here q^2 has only one zero where $z = z_0$, but it has an infinity where $z = z_1$. The curvature of the curve of q^2 versus z at $z = z_0$ is $-2\alpha/(z_0 - z_1)^2$. If this is to be small, $|z_0 - z_1|$ must be large, so that the infinity and the zero must be well separated. When the expression (8.2) is used for q^2, therefore, it implies that the zero of q at $z = z_0$ is at a great distance both from other zeros and from infinities. A method of solving the differential equations that allows for a small curvature of the function $q^2(z)$ near its zero is described in § 8.20.

8.3. The differential equation for horizontal polarisation and oblique incidence

The differential equations to be satisfied by the wave fields for oblique incidence were derived in § 7.2 for the case when the earth's magnetic field is neglected. It was shown that they separate into two sets, one for fields in which the electric vector is everywhere horizontal, and the other for fields in which the electric vector is everywhere in the plane of incidence. These two cases are usually said to apply to horizontal and vertical polarisation respectively. For horizontal polarisation the equations are (7.3) or (7.5) or (7.6), where q is given by (7.2). If the effect of electron

collisions is neglected, so that $Z = 0$, then (7.2) reduces to (8.1). If $S^2 = 0$ so that $C^2 = 1$, (7.2) gives $q^2 = n^2$, and the differential equations apply for vertical incidence. The solution for vertical incidence is therefore a special case of the solution for oblique incidence with horizontal polarisation.

When (8.2) is inserted in (7.6) it gives

$$\frac{d^2E_y}{dz^2} - k^2a(z - z_0)E_y = 0. \tag{8.5}$$

This is a very important differential equation which governs the behaviour of the wave fields near a zero of q. Its properties are given in the following sections.

For 'vertical' polarisation the differential equations are (7.4) or (7.68). When (8.2) is inserted in (7.4) it gives an equation which is more complicated than (8.5); the discussion of this case is postponed until §§ 15.5–15.7.

8.4. The Stokes differential equation

The theory in the rest of this chapter refers only to horizontal polarisation, so that E_y is the only non-zero component of the electric field. The subscript y will therefore be omitted. In (8.5) it is convenient to use the new independent variable

$$\zeta = (k^2a)^{\frac{1}{3}}(z - z_0), \tag{8.6}$$

where the value of $(k^2a)^{\frac{1}{3}}$ is taken to be real and positive. Thus ζ is a measure of the height. Then (8.5) becomes

$$\frac{d^2E}{d\zeta^2} = \zeta E. \tag{8.7}$$

This is known as the Stokes differential equation. The same name is sometimes given to the equation $d^2E/dx^2 + xE = 0$ which is easily converted to (8.7) by the substitution $x = (-1)^{\frac{1}{3}}\zeta$.

In this book the name Stokes is used for two distinct purposes. First, it is used for the differential equation (8.7) or its equivalent. Second, it is used for the Stokes phenomenon, § 8.12. This was discovered by Stokes originally for solutions of (8.7) but it can occur for any function with an irregular singularity. The terms 'Stokes lines' and 'anti-Stokes lines' and 'Stokes diagram' are used as part of the description of the Stokes phenomenon and are explained in §§ 8.13, 8.14. Many of the functions of theoretical physics have irregular singularities at infinity and there are examples in ch. 15. They display the Stokes phenomenon but they are not necessarily connected with solutions of (8.7).

Equation (8.7) has no singularities when ζ is bounded and its solution must therefore be bounded and single valued, except possibly at $\zeta = \infty$. It is necessary to study the properties of these solutions for both real and complex values of ζ.

Solutions of (8.7) can be found as series in ascending powers of ζ, by the standard method. Assume that a solution is $E = a_0 + a_1\zeta + a_2\zeta + \ldots$. Substitute this in (8.7)

and equate powers of ζ. This gives relations between the constants a_0, a_1, a_2, etc., and leads finally to

$$E = a_0\left\{1 + \frac{\zeta^3}{3\cdot2} + \frac{\zeta^6}{6\cdot5\cdot3\cdot2} + \frac{\zeta^9}{9\cdot8\cdot6\cdot5\cdot3\cdot2} + \ldots\right\} + a_1\left\{\zeta + \frac{\zeta^4}{4\cdot3} + \frac{\zeta^7}{7\cdot6\cdot4\cdot3} + \frac{\zeta^{10}}{10\cdot9\cdot7\cdot6\cdot4\cdot3} + \ldots\right\}, \tag{8.8}$$

which contains the two arbitrary constants a_0 and a_1, and is therefore the most general solution. The series are convergent for all ζ, which confirms that every solution of (8.7) is bounded, continuous and single valued. Series for the derivative $dE/d\zeta$ are easily found from (8.8). The two series (8.8) separately have no particular physical significance. The constants a_0 and a_1 for the functions $\mathrm{Ai}(\zeta)$ and $\mathrm{Bi}(\zeta)$ can be found from (8.16), (8.17) below or from their derivatives with respect to ζ, by putting $\zeta = 0$. Their values are as follows:

For $\mathrm{Ai}(\zeta)$:

$$a_0 = 3^{-\frac{2}{3}}/(-\tfrac{1}{3})! = 0.355\,03,$$
$$a_1 = -3^{-\frac{1}{3}}/(-\tfrac{2}{3})! = -0.258\,82. \tag{8.9}$$

For $\mathrm{Bi}(\zeta)$:

$$a_0 = 3^{-\frac{1}{6}}/(-\tfrac{1}{3})! = 0.614\,93,$$
$$a_1 = 3^{\frac{1}{6}}/(-\tfrac{2}{3})! = 0.448\,29.$$

These values, with the series (8.8) and their derivatives, may be used for computing for real and complex ζ up to about $|\zeta| = 5.0$, though in radio propagation problems they are rarely needed for values of $|\zeta|$ greater than about 2.0.

8.5. Qualitative discussion of the solutions of the Stokes equation

Equation (8.7) shows that, if E and ζ are real, $d^2E/d\zeta^2$ is real. If $dE/d\zeta$ is also real, than E must be real for all real values of ζ. This is also apparent from (8.8). It is of interest to trace the curve of E versus ζ when ζ and E are both real. If ζ is positive, (8.7) shows that the curvature of the curve has the same sign as E. Hence the curve is convex towards the line $E = 0$. If the curve is traced step by step from $\zeta = 0$ upwards, there are three possibilities which are illustrated in fig. 8.3. First, if the initial slope is sufficiently negative, the curve can cross the line $E = 0$ (curve A). When it does so, the sign of the curvature changes, and for higher values of ζ the magnitude of the slope must increase indefinitely. Hence the curve moves indefinitely further from the line $E = 0$ and can never cross it again. Secondly, if the initial slope is positive or only slightly negative, the slope can become zero before the curve reaches the line $E = 0$ (curve B). Thereafter the curve moves indefinitely further from this line and can never cross it. In both these cases E ultimately becomes indefinitely large as ζ increases. The third possibility occurs for one particular negative value of the initial slope. The curve then approaches the line $E = 0$, and never actually reaches it, but

gets closer and closer to it (curve C). The slope must always have the opposite sign to E, and E must become smaller and smaller as ζ increases. This last case is of particular importance, and the solution $\mathrm{Ai}(\zeta)$, described later, has this property.

When ζ is negative, the curvature has the opposite sign to E. Hence the curve is concave towards the line $E = 0$. If the curve is traced step by step from $\zeta = 0$ towards increasingly negative values of ζ, it must always curve towards the line $E = 0$ and eventually cross it. The sign of the curvature then changes so that the curve again bends towards the line $E = 0$, and crosses it again. In all cases therefore the function E is oscillatory and as ζ becomes more negative the curvature for a given E increases, so that the oscillation gets more rapid and its amplitude gets smaller. This is illustrated for all three curves in fig. 8.3.

8.6. Solutions of the Stokes equation expressed as contour integrals

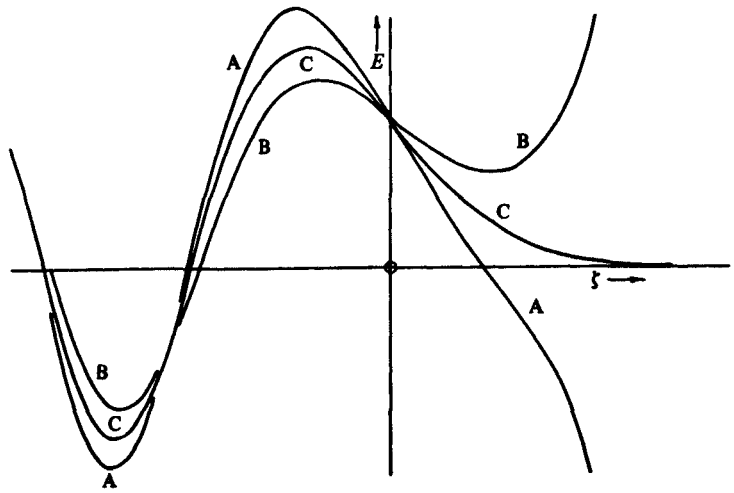
A useful form of the solutions of (8.7) can be found in the form of a contour integral. Let

$$E = \int_a^b \mathrm{e}^{\zeta t} h(t)\,\mathrm{d}t, \tag{8.10}$$

where t is a complex variable and the integral is evaluated along some path in the complex t plane, whose end points a and b are to be specified later. An integral of the form (8.10) is said to be of Laplace type. This form can always be used for a linear differential equation whose coefficients are linear functions of the independent variable. Since (8.10) must satisfy (8.7), it is necessary that

$$\int_a^b (t^2 - \zeta) h(t) \mathrm{e}^{\zeta t}\,\mathrm{d}t = 0. \tag{8.11}$$

Fig. 8.3. Behaviour of solutions $E(\zeta)$ of the Stokes equation.



The second term can be integrated by parts, which gives

$$-\mathrm{e}^{\zeta t}h(t)\Big|_a^b + \int_a^b \left\{t^2 h(t) + \frac{\mathrm{d}h(t)}{\mathrm{d}t}\right\}\mathrm{e}^{\zeta t}\,\mathrm{d}t = 0. \tag{8.12}$$

The limits a, b are to be chosen so that the first term vanishes at both limits. Then (8.11) is satisfied if

$$\frac{\mathrm{d}h(t)}{\mathrm{d}t} + t^2 h(t) = 0, \tag{8.13}$$

that is if

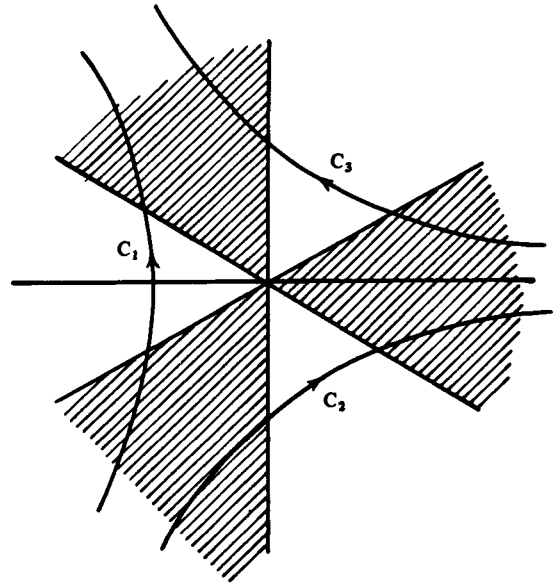
$$h(t) = A\exp(-\tfrac{1}{3}t^3) \tag{8.14}$$

where A is a constant. The limits a and b must therefore be chosen so that $\exp(-\frac{1}{3}t^3 + \zeta t)$ is zero for both. This is only possible if $|t| \to \infty$ and $2\pi r - \frac{1}{2}\pi < 3\arg t < 2\pi r + \frac{1}{2}\pi$, where r is an integer. Fig. 8.4 is a diagram of the complex t plane, and a and b must each be at infinity in one of the shaded sectors. They cannot both be in the same sector, for then the integral (8.10) would be zero. Hence the path may be chosen in three ways, as shown by the three curves $\mathrm{C}_1, \mathrm{C}_2, \mathrm{C}_3$. This might appear at first to give three independent solutions of (8.7). But the path C_1 can be distorted so as to coincide with the two paths $\mathrm{C}_2 + \mathrm{C}_3$, so that

$$\int_{\mathrm{C}_1} = \int_{\mathrm{C}_2} + \int_{\mathrm{C}_3}, \tag{8.15}$$

and therefore there are only two independent solutions.

Fig. 8.4. The complex t plane, with possible paths for the integral representation (8.10) of solutions of the Stokes equation.



Jeffreys and Jeffreys (1972) define the two functions $\mathrm{Ai}(\zeta)$ and $\mathrm{Bi}(\zeta)$ as follows:

$$\mathrm{Ai}(\zeta) = \frac{1}{2\pi \mathrm{i}} \int_{C_1} \exp(-\tfrac{1}{3}t^3 + \zeta t)\,\mathrm{d}t, \tag{8.16}$$

$$\mathrm{Bi}(\zeta) = \frac{1}{2\pi} \int_{C_2} \exp(-\tfrac{1}{3}t^3 + \zeta t)\,\mathrm{d}t - \frac{1}{2\pi} \int_{C_3} \exp(-\tfrac{1}{3}t^3 + \zeta t)\,\mathrm{d}t. \tag{8.17}$$

In (8.16) the path C_1 can be distorted so as to coincide with the imaginary t axis for almost its whole length. It must be displaced very slightly to the left of this axis at its ends. Let $t = \mathrm{i}s$. Then (8.16) becomes

$$\mathrm{Ai}(\zeta) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\{\mathrm{i}(\zeta s + \tfrac{1}{3}s^3)\}\,\mathrm{d}s. \tag{8.18}$$

Here the imaginary part of the integrand is an odd function of s, and contributes nothing to the integral, which may therefore be written

$$\mathrm{Ai}(\zeta) = \frac{1}{\pi} \int_{0}^{\infty} \cos(\zeta s + \tfrac{1}{3}s^3)\,\mathrm{d}s. \tag{8.19}$$

Apart from a constant, this is the same as the expression used by Airy (1838, 1849). It is known as the Airy integral, and $\mathrm{Ai}(\zeta)$ is called the Airy integral function.

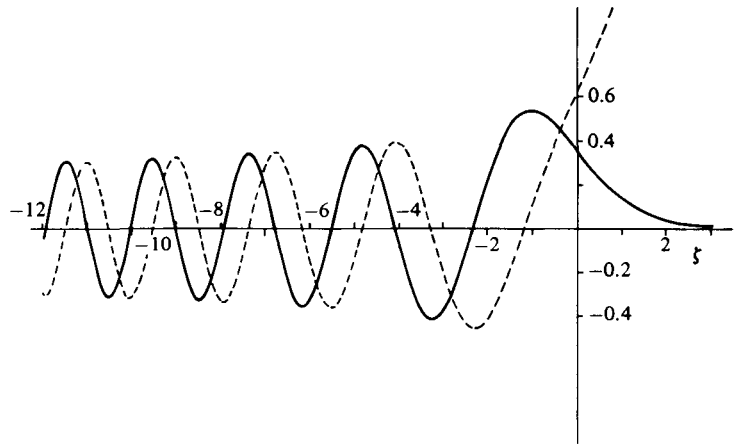
The functions $\mathrm{Ai}(\zeta)$ and $\mathrm{Bi}(\zeta)$ are shown in fig. 8.5.

8.7. Solutions of the Stokes equation expressed as Bessel functions

The Stokes equation (8.7) can be converted into Bessel's equation by changing both the dependent and independent variables. First let

$$\tfrac{2}{3}\zeta^{\frac{3}{2}} = \mathrm{i}\xi. \tag{8.20}$$

Fig. 8.5. The functions $\mathrm{Ai}(\zeta)$ continuous curve, and $\mathrm{Bi}(\zeta)$ broken curve.



Then (8.7) becomes

$$\frac{\mathrm{d}^2E}{\mathrm{d}\xi^2}+\frac{1}{3\xi}\frac{\delta E}{\mathrm{d}\xi}+E=0. \tag{8.21}$$

Next let

$$E=\xi^{\frac{1}{2}}\mathscr{E}=(-\tfrac{2}{3}\mathrm{i})^{\frac{1}{3}}\zeta^{\frac{1}{2}}\mathscr{E}. \tag{8.22}$$

Then (8.21) becomes

$$\frac{\mathrm{d}^2\mathscr{E}}{\mathrm{d}\xi^2}+\frac{1}{\xi}\frac{\mathrm{d}\mathscr{E}}{\mathrm{d}\xi}+\left(1-\frac{1}{9\xi^2}\right)\mathscr{E}=0. \tag{8.23}$$

This is Bessel's equation of order one-third. The transformations (8.20) and (8.22) are complicated and introduce ambiguities. For example, (8.20) does not define ξ completely, since there is an ambiguity of the sign of the term on the left. Consequently the use of the Bessel function solutions can lead to errors. Moreover, one solution of Bessel's equation (8.23) has a singularity where $\xi=0$, whereas the solutions of the Stokes equation have no singularities. The transformations (8.20) and (8.22) in fact introduce compensating singularities. The relations between the Bessel functions and Ai(ζ), Bi(ζ) are given by Watson (1944), Miller (1946), Olver (1974), Abramowitz and Stegun (1965).

8.8. Tables of the Airy integral functions. Computing

Many tables of the functions Ai(ζ), Bi(ζ) and related functions are available, and a list is given by Abramowitz and Stegun (1965). These are nearly all for real ζ only. One of the most useful is that of Miller (1946) who also gives a summary of the properties of these functions.

Tables for complex ζ are given by Woodward and Woodward (1946). Tables of the functions

$$\begin{aligned}h_1(z)&=(12)^{\frac{1}{6}}\exp(-\tfrac{1}{6}\pi\mathrm{i})\{\mathrm{Ai}(-z)-\mathrm{iBi}(-z)\}\\ h_2(z)&=(12)^{\frac{1}{6}}\exp(\tfrac{1}{6}\pi\mathrm{i})\{\mathrm{Ai}(-z)+\mathrm{iBi}(-z)\}\end{aligned} \tag{8.24}$$

and their derivatives, for complex z, are given by the staff of the Computation Laboratory at Cambridge, Mass. (1945). From them Ai(z) and Bi(z) and their derivatives can be found.

Subroutines for computing Ai(ζ), Bi(ζ) and their derivatives are available in the libraries of many of the larger computers, although in some cases only for real values of ζ. For real or complex ζ the series (8.8) with (8.9) can be used for $|\zeta|$ less than about 4 or 5. For larger $|\zeta|$ the asymptotic forms (8.52)–(8.61) below can be used, and, if greater accuracy is needed, these forms can be combined with asymptotic expansions (8.30)–(8.32).

8.9. Zeros and turning points of Ai(ζ) and Bi(ζ)

From the discussion of §(8.5) it is clear that Ai(ζ), Ai′(ζ), Bi(ζ) and Bi′(ζ) have no zeros when ζ is real and positive. Each of them has an infinite sequence of zeros when ζ is

real and negative; compare fig. 8.5. It can be shown that $\mathrm{Ai}(\zeta)$ and $\mathrm{Ai}'(\zeta)$ have no other zeros, but $\mathrm{Bi}(\zeta)$ and $\mathrm{Bi}'(\zeta)$ have an infinite sequence of zeros on each of the lines $\arg\zeta = \pm\pi/3$. The proof of these results is given by Olver (1974, §§ 11.5, 11.8).

8.10. The W.K.B. solutions of the Stokes equation

Approximate solutions of the Stokes equation (8.7) can be found by the W.K.B. method of § 7.6. Thus in (8.7) let

$$E = A\exp\{\mathrm{i}\phi(\zeta)\}. \tag{8.25}$$

Then the method gives (compare (7.18), (7.19)):

$$\left(\frac{\mathrm{d}\phi}{\mathrm{d}\zeta}\right)^2 = -\zeta + \mathrm{i}\frac{\mathrm{d}^2\phi}{\mathrm{d}\zeta^2}. \tag{8.26}$$

The steps of (7.19)–(7.26) are now followed through and lead to the two W.K.B. solutions

$$E = \zeta^{-\frac{1}{4}}\exp(-\tfrac{2}{3}\zeta^{\frac{3}{2}}) \tag{8.27}$$

and

$$E = \zeta^{-\frac{1}{4}}\exp(\tfrac{2}{3}\zeta^{\frac{3}{2}}). \tag{8.28}$$

For large enough $|\zeta|$ any solution of (8.7) is a linear combination of (8.27), (8.28) but not the same combination for all $\arg\zeta$; see § 8.12. The combinations for $\mathrm{Ai}(\zeta)$, $\mathrm{Bi}(\zeta)$ are given later at (8.52), (8.53) and (8.56)–(8.58) respectively. A condition that these are good approximations is that $|\zeta|$ shall be large enough. It is useful to study them when $|\zeta| = 1$. The function $\mathrm{Ai}(\zeta)$ has only the one term (8.27) in its asymptotic approximation when $0 < \arg\zeta < \frac{2}{3}\pi$, and this range runs from one Stokes line to the next; see § 8.13. The function $\mathrm{Ai}(\zeta)$ and its asymptotic approximation from (8.52) were computed and compared. The minimum error was 7.0% for $\arg\zeta = 0$ where the term is subdominant. The maximum error was 14.3% for $\arg\zeta = \frac{2}{3}\pi$ where the term is dominant. On the anti-Stokes line, $\arg\zeta = \frac{1}{3}\pi$, the error was 8.2%. The average error over this whole range was 8.9%. For $|\zeta| = 2$ the corresponding figures were 3.1%, 5.3%, 3.4%, 3.7%. These results suggest that if

$$|\zeta| \geqslant 1, \tag{8.29}$$

the error in using the asymptotic forms alone is less than about 8 to 9%. This criterion is used as a rough guide in this book.

8.11. Asymptotic expansions

In many books on linear differential equations there is a discussion of the properties of the solutions at various points in the complex plane of the independent variable. The points are classified into (a) ordinary points, (b) regular singular points or regular singularities, and (c) irregular singular points or irregular singularities.

Regular singularities are not used in this section but are mentioned later, §§ 15.5, 15.12 ff. Every differential equation must have regular or irregular singular points and for nearly all the differential equations of physics it is found that the point at infinity is an irregular singularity. An exception is the hypergeometric equation, discussed in § 15.12 ff. At an ordinary point, say $\zeta = \zeta_0$, it is always possible to find two convergent series solutions of the differential equation in ascending integer powers of $\zeta - \zeta_0$, and the radius of convergence is the distance from ζ_0 to the nearest singular point. For the Stokes equation (8.7) every point of the complex ζ plane except infinity is an ordinary point. The series (8.8) are the convergent expansions about the origin $\zeta = 0$, and since the only singular point is at infinity, they are convergent for all ζ.

For the region near an irregular singularity the series method can sometimes be used, but the series is divergent and might not use integer powers. This applies to the Stokes equation, as shown below, and for most other irregular singularities that arise in physical problems. A divergent series cannot define any function but it can still be used to give important properties of the solution. Thus we have to deal with two important things: (i) solutions near an irregular singularity can be studied by using divergent series; (ii) the most commonly occurring irregular singularity is at infinity. These two ideas are really quite separate. The fact that they usually occur together can sometimes lead to confusion.

To study the behaviour at infinity it is sometimes convenient to change the independent variable to $x = 1/\zeta$ so that the point $\zeta = \infty$ is transformed to the origin of the complex x plane, and we study the resulting differential equation near its irregular singularity at $x = 0$. Then by using a trial series in ascending powers of x and substituting in the differential equation, we can formally find the coefficients but the series is found to be divergent. There is, however, no need to make this transformation. The series is the same as a series in descending powers of ζ. We can use such a trial series in the original differential equation.

For the Stokes equation (8.7), the W.K.B. solutions (8.27), (8.28) are good approximations, according to (8.29), when $|\zeta|$ is large. But these functions cannot be expanded in descending powers of ζ, and indeed it is not possible to find a solution of (8.7) in descending powers of ζ. But we may take as a trial solution

$$E \sim \zeta^{-\frac{1}{4}} \exp(-\tfrac{2}{3}\zeta^{\frac{3}{2}})\{1 + a_1\zeta^{-r} + a_2\zeta^{-2r} + \cdots\} \tag{8.30}$$

where r is positive. If this is substituted in (8.7), a factor $\exp(-\frac{2}{3}\zeta^{\frac{3}{2}})$ may be cancelled from every term. If then the coefficients of the various powers of ζ in succession are equated to zero, it is found that $r = \frac{3}{2}$ and the coefficients $a_1, a_2 \cdots$ can be found. Similarly the trial solution

$$E \sim \zeta^{-\frac{1}{4}} \exp(\tfrac{2}{3}\zeta^{\frac{3}{2}})\{1 + b_1\zeta^{-r} + b_2\zeta^{-2r} + \cdots\} \tag{8.31}$$

may be used. The same process again yields $r = \frac{3}{2}$, and b_1, b_2 can be found. The

values are

$$b_m = (-1)^m a_m = \frac{(3m-\frac{1}{2})!}{(2m)!(-\frac{1}{2})!3^{2m}} = \frac{(6m-1)(6m-3)\cdots(2m+1)}{m!(144)^m}. \tag{8.32}$$

The standard convergence tests show that the series in (8.30), (8.31) are divergent for all values of ζ.

These are examples of the type of series known as 'asymptotic expansions'. Many physicists find that this is a difficult subject, and the reader who wants to master it must expect to devote much effort and time to it. For accounts of it see Olver (1974), Jeffreys and Jeffreys (1972), Jeffreys (1962), Dingle (1973). Here we give only a brief summary of how these expansions are used.

To avoid repeatedly writing the fractional power $r = \frac{3}{2}$, let

$$s = \tfrac{2}{3}\zeta^{\frac{3}{2}}. \tag{8.33}$$

Then the series representation (8.30) may be written

$$E\zeta^{\frac{1}{4}}e^s \equiv Y(s) \sim 1 + c_1/s + c_2/s^2 + \cdots + c_m/s^m + \cdots. \tag{8.34}$$

To study the function $Y(s)$ we must have some definition of it. It is often expressible as a contour integral. In the present example, E in (8.34) is proportional to (8.16). If the series in (8.34) does not converge it does not define any number and so the = sign cannot be used. We use the sign $\sim$ whose meaning is explained below. Even if (8.34) is divergent, it can always be made to 'converge at first' by making $|s|$ large enough. For example we can ensure that the terms decrease steadily as m increases up to the N^{th} term, by taking $|s| > Q$ where Q is the greatest value of $|c_{m+1}/c_m|$ for $m < N$. If $|s|$ is made extremely large, some of the terms after the first will be extremely small. In series such as (8.31), however, the much later terms get large no matter how large $|s|$ is. But it is clear that there must be a smallest term. Which numbered term it is depends on $|s|$.

Let $S_m(s)$ denote the sum of the first $m+1$ terms of the series

$$S_m(s) = 1 + c_1/s + \cdots + c_m s^m. \tag{8.35}$$

Then the difference

$$R_m(s) = Y(s) - S_m(s) \tag{8.36}$$

is called the 'remainder'. For fixed s, $|R_m(s)|$ will not get small when m gets large because the series is divergent. But it is possible that, for fixed m, $|R_m(s)|$ gets small when $|s|$ gets large. This leads to the Poincaré definition. If for every fixed value of m

$$s^m R_m(s) \to 0 \text{ as } |s| \to \infty \tag{8.37}$$

then we say that the series (8.34) is an asymptotic expansion for $Y(s)$. This gives the meaning of the $\sim$ sign.

Note that the Poincaré definition is concerned only with what happens when $|s| \to \infty$. It is not concerned with non-infinite values of s.

To use a divergent asymptotic expansion the series must be truncated in some

way. It can be shown that the error is least when the series is stopped at or just before the smallest term. Then the modulus of the remainder is less than the last retained term of the series.

If $|\zeta|$ is large enough in (8.30), many terms of the series after the first are small. In many practical problems, including most of those discussed in this book, these terms are neglected and we use only the first term. The resulting expression is called an 'asymptotic form'. All W.K.B. solutions are asymptotic forms in this sense. The series and the Poincaré definition are needed for reference in the discussions of later sections. Occasionally it is useful to use a few terms of the asymptotic series when the W.K.B. solutions alone are not accurate enough.

8.12. The Stokes phenomenon of the 'discontinuity of the constants'

The expressions (8.27) and (8.28) are multiple valued functions, whereas any solution of (8.7) is single valued. Hence a solution of (8.7) cannot be represented by the same combination of (8.27) and (8.28) for all values of ζ. To illustrate this, consider the function $\mathrm{Ai}(\zeta)$. It was mentioned in § 8.5 that when ζ is real and positive, this function decreases steadily as ζ increases (fig. 8.3, curve C), but never becomes zero. Hence the W.K.B. approximation for $\mathrm{Ai}(\zeta)$ when ζ is real and positive must be given by

$$\mathrm{Ai}(\zeta) \sim A\zeta^{-\frac{1}{4}} \exp(-\tfrac{2}{3}\zeta^{\frac{3}{2}}) \quad (\text{for } \arg\zeta = 0), \tag{8.38}$$

where A is a constant and the positive values of $\zeta^{-\frac{1}{4}}$, $\zeta^{\frac{3}{2}}$ are used. It cannot include a multiple of (8.28) for this would make the function increase indefinitely for large ζ. Now let $|\zeta|$ be kept constant, and let $\arg\zeta$ increase continuously from 0 to π, so that ζ becomes real and negative. Then the expression (8.38) becomes

$$A\mathrm{e}^{-\mathrm{i}\frac{1}{4}\pi}|\zeta^{-\frac{1}{4}}|\exp\{\tfrac{2}{3}\mathrm{i}|\zeta^{\frac{3}{2}}|\} \quad (\text{for } \arg\zeta = \pi), \tag{8.39}$$

which is a complex function. But $\mathrm{Ai}(\zeta)$ is real for all real values of ζ, so that (8.39) cannot represent $\mathrm{Ai}(\zeta)$ when ζ is real and negative, and the correct representation must include multiples of both (8.27) and (8.28). It is shown later that an additional term should be added to (8.38) when $\arg\zeta > \frac{2}{3}\pi$.

The W.K.B. approximation to the most general solution of the Stokes equation is

$$A\zeta^{-\frac{1}{4}} \exp(-\tfrac{2}{3}\zeta^{\frac{3}{2}}) + B\zeta^{-\frac{1}{4}} \exp(\tfrac{2}{3}\zeta^{\frac{3}{2}}), \tag{8.40}$$

but this can apply only to a part of the complex ζ plane. If ζ moves out of this part, one of the arbitrary constants A or B must be changed. This phenomenon was discovered by Stokes (1858), and is called the 'Stokes phenomenon of the discontinuity of the arbitrary constants.' See the note in § 8.4 about the two distinct uses of the name Stokes in this book.

8.13. Stokes lines and anti-Stokes lines

The exponents of both the exponentials in (8.40) are real if

$$\arg\zeta = 0, \tfrac{2}{3}\pi, \text{ or } \tfrac{4}{3}\pi, \tag{8.41}$$

and then if $|\zeta|$ becomes indefinitely large, one exponential becomes indefinitely small, and the other indefinitely large. Moreover, if $|\zeta|$ is kept constant, and $\arg\zeta$ is varied, both exponentials have maximum or minimum values when $\arg\zeta$ is given by (8.41), which defines three lines radiating from the origin of the complex ζ plane. These are known as the 'Stokes lines'.

The exponents have equal moduli if

$$\arg\zeta = \tfrac{1}{3}\pi, \pi, \text{ or } \tfrac{5}{3}\pi, \tag{8.42}$$

and then if $|\zeta|$ becomes indefinitely large, the moduli of both exponentials remain equal to unity, but the terms oscillate more and more rapidly. The radial lines defined by (8.42) are called the 'anti-Stokes lines'. This nomenclature is used here and is found in many papers on radio propagation. It is not accepted by all authors, however, and for example Olver (1974, p. 518) recommends a different nomenclature.

For all values of $\arg\zeta$ except those in (8.42) one exponential must have modulus greater than unity, and the other must have modulus less than unity. The larger exponential remains so, as long as ζ lies in the 120° sector between two anti-Stokes lines, and the corresponding term in (8.40) is called the 'dominant' term. The other term, containing the smaller exponential, is called the 'subdominant' term. Each term changes from dominant to subdominant, or the reverse, when ζ crosses an anti-Stokes line.

It was shown in § 8.11 that one of the constants A and B must change when ζ crosses some line in the complex ζ plane. It is fairly clear that the constant in the dominant term cannot change, for this would give a detectable discontinuity in the function (8.40), which would mean that it would not even approximately satisfy the differential equation. Hence the constant which changes must be that in the subdominant term, and the most likely place for the change to occur is on a Stokes line, for there the ratio of the subdominant to the dominant term is smallest.

The solution (8.40) is approximate. The two terms are W.K.B. solutions and both are subject to error. To get a more accurate solution, these terms may each be multiplied by a divergent asymptotic expansion, § 8.11. To get the greatest accuracy this must be truncated just before or just after the smallest term. This smallest term then gives an upper bound for the error. It was shown by Stokes (1858) that when the multiplier of the subdominant term changes on a Stokes line, the resulting discontinuity is less than the error bound for the dominant term.

In the following two sections it will be assumed that the arbitrary constant in the subdominant term may change on a Stokes line. These sections are descriptive and are intended to help towards an understanding of the Stokes phenomenon. A more formal mathematical proof of the results is given in § 9.7.

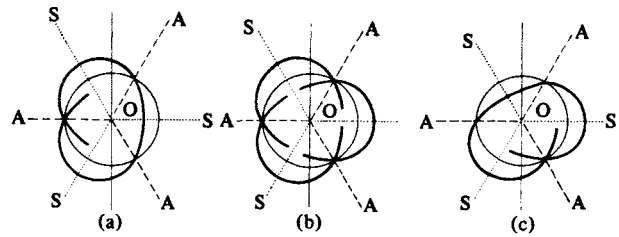
8.14. The Stokes diagram

In any 120° sector between two Stokes lines, a solution of the Stokes equation may contain either one or both of the terms in (8.40). To indicate the nature of the W.K.B. or asymptotic approximation to a particular solution, Stokes (1858) used a diagram constructed as follows. Radial lines are drawn from a fixed point O (fig. 8.6) in the directions of the Stokes and anti-Stokes lines, and labelled S or A respectively. A circle is drawn with centre O. In this diagram radial directions indicate values of $\arg\zeta$ but radial distances do not indicate values of $|\zeta|$. The plane of the diagram, therefore, is not the complex ζ plane. One or two thick lines are drawn in the diagram, inside and outside the circle, and they cross the circle on the anti-Stokes lines. For some value of $\arg\zeta$ a radial line is imagined to be drawn in the corresponding direction. If it crosses a thick line inside the circle, this shows that there is a subdominant term in the asymptotic approximation. If it crosses a thick line outside the circle, this shows that there is a dominant term.

Where the thick line crosses a Stokes line inside the circle, the constant multiplying the associated subdominant term may change, and this is indicated by a break in the thick line. The thick line crosses the circle on the anti-Stokes lines because there the associated terms change from dominant to subdominant or the reverse. The thick line must remain unbroken except where it meets a Stokes line inside the circle, for only there can the associated constant change. Moreover, this change can only occur if the dominant term is present. If it were absent, a change in the constant of the subdominant term would give a detectable discontinuity, since there is now no dominant term to mask it.

These properties are illustrated in fig. 8.6(a), which is the Stokes diagram for the function Ai(ζ). In the sector $-\frac{2}{3}\pi < \arg\zeta < \frac{2}{3}\pi$ there is only one term in the asymptotic approximation, and this is subdominant in the sector $-\frac{1}{3}\pi < \arg\zeta < \frac{1}{3}\pi$. The constant cannot change on the Stokes line at $\arg\zeta = 0$ because there is no dominant term. The same term is dominant on the Stokes line at $\arg\zeta = \frac{2}{3}\pi$, and for greater values of $\arg\zeta$ there is also a subdominant term, which becomes dominant

Fig. 8.6. Stokes diagrams for the functions (a) Ai(ζ), (b) Bi(ζ), (c) $\mathrm{Ai}\{\zeta\exp(\frac{4}{3}\mathrm{i}\pi)\} = \mathrm{Ai}\{\zeta\exp(-\frac{2}{3}\mathrm{i}\pi)\}$.



when $\arg\zeta > \pi$, and here the original term again becomes subdominant. This term disappears beyond the Stokes line at $\arg\zeta = \frac{4}{3}\pi$ or $-\frac{2}{3}\pi$.

Another example is given in fig. 8.6(b), which is the Stokes diagram for the function $\mathrm{Bi}(\zeta)$. Here there are both dominant and subdominant terms for all values of $\arg\zeta$, and the constant in the subdominant term changes on all three Stokes lines.

8.15. Definition of the Stokes multiplier

It is now necessary to determine by how much the constant in the subdominant term changes when a Stokes line is crossed. Suppose that in the sector $0 < \arg\zeta < \frac{2}{3}\pi$ (sector I of fig. 8.7) a given solution of the Stokes equation has the asymptotic approximation

$$\zeta^{-\frac{1}{4}}[A_1 \exp(-\tfrac{2}{3}\zeta^{\frac{3}{2}}) + B_1 \exp(\tfrac{2}{3}\zeta^{\frac{3}{2}})]. \tag{8.43}$$

On the Stokes line at $\arg\zeta = \frac{2}{3}\pi$ (S_2 in fig. 8.7) the first term is dominant, and hence for the sector $\frac{2}{3}\pi < \arg\zeta < \frac{4}{3}\pi$ (sector II) the asymptotic approximation is

$$\zeta^{-\frac{1}{4}}[A_1 \exp(-\tfrac{2}{3}\zeta^{\frac{3}{2}}) + B_2 \exp(\tfrac{2}{3}\zeta^{\frac{3}{2}})]. \tag{8.44}$$

The constant in the subdominant term has changed by $B_2 - B_1$. Now this change is zero if A_1 is zero. It cannot depend on B_1, for it would be unaltered if we added to (8.43) any multiple of the solution in which $A_1 = 0$. Since the differential equation is linear, $B_2 - B_1$ must be proportional to A_1, so that

$$B_2 - B_1 = \lambda_2 A_1, \tag{8.45}$$

where λ_2 is a constant called the 'Stokes multiplier' for the Stokes line at $\arg\zeta = \frac{2}{3}\pi$. It gives the change in the constant for the subdominant term when the Stokes line is crossed in an anticlockwise direction. If the crossing is clockwise, the Stokes multiplier has the opposite sign. Stokes multipliers can be defined in a similar way for the other two Stokes lines. It will be shown in §§ 8.16, 9.7, that for the Stokes equation (8.7) all three Stokes multipliers are equal to i.

8.16. Furry's derivation of the Stokes multipliers for the Stokes equation

The following derivation of the Stokes multipliers seems to have been used first by Furry (1947). The two terms in (8.43) are multiple valued functions of ζ with a branch point at the origin. Hence we introduce a cut in the complex ζ plane from 0 to $-\infty$ along the real axis, and take $-\pi \leqslant \arg\zeta \leqslant \pi$. Consider a solution whose asymptotic approximation is (8.43) in sector I of fig. 8.7. Then in the top part of sector II it is

$$\zeta^{-\frac{1}{4}}[A_1 \exp(-\tfrac{2}{3}\zeta^{\frac{3}{2}}) + (B_1 + \lambda_2 A_1)\exp(\tfrac{2}{3}\zeta^{\frac{3}{2}})]. \tag{8.46}$$

On the Stokes line S_1 the second term of (8.43) is dominant, and the constant in the first term changes so that in sector III the asymptotic approximation is

$$\zeta^{-\frac{1}{4}}[(A_1 - \lambda_1 B_1)\exp(-\tfrac{2}{3}\zeta^{\frac{3}{2}}) + B_1 \exp(\tfrac{2}{3}\zeta^{\frac{3}{2}})], \tag{8.47}$$

where λ_1 is the Stokes multiplier for the Stokes line S_1. On the Stokes line S_3 the first term of (8.47) is dominant, and the constant in the second term changes so that in the lower part of sector II the asymptotic approximation is

$$\zeta^{-\frac{1}{4}}[(A_1 - \lambda_1 B_1)\exp(-\tfrac{2}{3}\zeta^{\frac{3}{2}}) + \{B_1 - \lambda_3(A_1 - \lambda_1 B_1)\}\exp(\tfrac{2}{3}\zeta^{\frac{3}{2}})], \tag{8.48}$$

where λ_3 is the Stokes multiplier for the Stokes line S_3. Now the cut passes through the middle of sector II. The solution must be continuous across the cut, and hence (8.48) and (8.46) must agree. On crossing the cut from top to bottom, $\zeta^{-\frac{1}{4}}$ changes by a factor $e^{\frac{1}{2}i\pi}$, and $\zeta^{\frac{3}{2}}$ changes sign. Hence by equating coefficients of the exponentials in (8.46) and (8.48) we obtain

$$i(A_1 - \lambda_1 B_1) = B_1 + \lambda_2 A_1, \tag{8.49}$$

$$i\{B_1 - \lambda_3(A_1 - \lambda_1 B_1)\} = A_1. \tag{8.50}$$

Now this argument must apply whatever the values of A_1 and B_1. If $A_1 = 0$, (8.49) shows that $\lambda_1 = i$. If $B_1 = 0$, (8.49) gives $\lambda_2 = i$, and (8.50) gives $\lambda_3 = i$. Hence

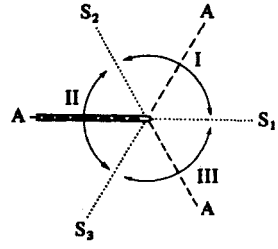
$$\lambda_1 = \lambda_2 = \lambda_3 = i. \tag{8.51}$$

From the differential equation (8.7) it could have been predicted that all three Stokes multipliers are the same, for this equation is unaltered if ζ is replaced by $\zeta e^{\frac{2}{3}\pi i}$, and it therefore has the same properties on each Stokes line.

8.17. The range of validity of asymptotic approximations

The discussion in §§ 8.12–8.16, and the Stokes diagram fig. 8.6 (a) for the function Ai(ζ), show that for the range $\frac{2}{3}\pi \leqslant \arg\zeta \leqslant \frac{4}{3}\pi$, the asymptotic approximation for Ai(ζ) has multiples of the two terms (8.27), (8.28) that contain the exponential factors $\exp(\pm\frac{2}{3}\zeta^{\frac{3}{2}})$. For all $\arg\zeta$ in this range, except the anti-Stokes line $\arg\zeta = \pi$, one of the two terms is dominant and the other is subdominant. For a fixed $\arg\zeta \neq \pi$ in this range, let $|\zeta|$ increase indefinitely. Then the subdominant term tends to zero and the ratio of Ai(ζ) to the dominant term tends to unity as $|\zeta|$ tends to ∞. Hence the asymptotic approximation to Ai(ζ) as given by the Poincaré definition of §8.11 need only contain the dominant term. It is only on the anti-Stokes line, where $\arg\zeta = \pi$ exactly, that it is necessary to include both terms, according to the Poincaré criterion.

Fig. 8.7. The complex ζ plane.



The Poincaré definition thus applies only to the limiting behaviour when $|\zeta| \to \infty$. When $|\zeta|$ is finite, it may clearly be insufficient to include only the dominant term. For example if $\arg\zeta = \pi$ the two terms are equal. If $|\zeta|$ is held constant and $\arg\zeta$ is made to differ slightly from π, the dominant term begins to increase and the subdominant term to decrease but at first the two terms are nearly equal, and in computing $\mathrm{Ai}(\zeta)$ it would be necessary to include both. But the Poincaré definition allows the subdominant term to be omitted. Care is therefore needed when specifying the range of $\arg\zeta$ over which a given asymptotic approximation is valid. For $\mathrm{Ai}(\zeta)$ the best way of doing this is as follows:

$$\mathrm{Ai}(\zeta) \sim \tfrac{1}{2}\pi^{-\frac{1}{2}}\zeta^{-\frac{1}{4}}\exp(-\tfrac{2}{3}\zeta^{\frac{3}{2}}) \quad \text{for} \quad -\tfrac{2}{3}\pi \leqslant \arg\zeta \leqslant \tfrac{2}{3}\pi, \tag{8.52}$$

$$\mathrm{Ai}(\zeta) \sim \tfrac{1}{2}\pi^{-\frac{1}{2}}\zeta^{-\frac{1}{4}}\{\exp(-\tfrac{2}{3}\zeta^{\frac{3}{2}}) + \mathrm{i}\exp(\tfrac{2}{3}\zeta^{\frac{3}{2}})\} \quad \text{for} \quad \tfrac{2}{3}\pi \leqslant \arg\zeta \leqslant \tfrac{4}{3}\pi, \tag{8.53}$$

where a fractional power ζ^f always means $|\zeta^f|\exp\{\mathrm{i}f\arg\zeta\}$. The constant $\frac{1}{2}\pi^{-\frac{1}{2}}$ is the same as A in (8.38). It is found from the contour integral (8.16) which is evaluated by the method of steepest descents as explained in § 9.7. In (8.52), (8.53) the ranges of $\arg\zeta$ end on Stokes lines, and this convention is used in the present book. According to the Poincaré definition the range in (8.52) could be written $-\pi < \arg\zeta < \pi$, and in (8.53) it could be written $\frac{1}{3}\pi < \arg\zeta < \frac{5}{3}\pi$. These ranges end just short of anti-Stokes lines, and the inequality signs are $<$ and not $\leqslant$, so that the two anti-Stokes lines are just outside the range at the ends. The two ranges of $\arg\zeta$ used in (8.52), (8.53) do not overlap. The larger ranges allowed by the Poincaré definition do overlap. They have common parts where both dominant and subdominant terms are present. In the Poincaré sense it does not then matter whether the subdominant term is included or not.

In most textbooks of mathematics that give asymptotic approximations, the range of $\arg\zeta$ is assessed according to the Poincaré definition, and this can sometimes lead to errors. An example of this is given in § 15.10, for the parabolic cylinder function. There are some further comments on this topic in § 15.8.

The asymptotic approximations for $\mathrm{Ai}'(\zeta)$ will also be needed later. They can be found from those for $\mathrm{Ai}(\zeta)$ by differentiating the asymptotic expansion, which gives new expansions for $\mathrm{Ai}'(\zeta)$. The first terms of these are as follows:

$$\mathrm{Ai}'(\zeta) \sim -\tfrac{1}{2}\pi^{-\frac{1}{2}}\zeta^{\frac{1}{4}}\exp(-\tfrac{2}{3}\zeta^{\frac{3}{2}}) \quad \text{for} \quad -\tfrac{2}{3}\pi \leqslant \arg\zeta \leqslant \tfrac{2}{3}, \tag{8.54}$$

$$\mathrm{Ai}'(\zeta) \sim \tfrac{1}{2}\pi^{-\frac{1}{2}}\zeta^{\frac{1}{4}}\{-\exp(-\tfrac{2}{3}\zeta^{\frac{3}{2}}) + \mathrm{i}\exp(\tfrac{2}{3}\zeta^{\frac{3}{2}})\} \quad \text{for} \quad \tfrac{2}{3}\pi \leqslant \arg\zeta \leqslant \tfrac{4}{3}\pi. \tag{8.55}$$

8.18. The choice of a fundamental system of solutions of the Stokes equation

Since the Stokes differential equation (8.7) is of the second order, it has two independent solutions which may be chosen in various ways, and these may be called the fundamental solutions. Any other solution is then a linear combination of them. There is no absolute criterion for these solutions and their choice is a matter of

convenience. An obvious choice for one fundamental solution is the Airy integral function Ai(ζ). As a second solution Jeffreys and Jeffreys (1972) use the function Bi(ζ) defined in § 8.6. The choice of Ai(ζ) and Bi(ζ) has the advantage that both functions are real when ζ is real. The asymptotic approximation for Bi(ζ) may be found from the contour integrals (8.17) which are evaluated by the method of steepest descents as explained in § 9.7. The result is:

$$\mathrm{Bi}(\zeta) \sim \tfrac{1}{2}\pi^{-\frac{1}{2}}\zeta^{-\frac{1}{4}}\{\mathrm{i}\exp(-\tfrac{2}{3}\zeta^{\frac{3}{2}}) + 2\exp(\tfrac{2}{3}\zeta^{\frac{3}{2}})\} \text{ for } 0 \leqslant \arg\zeta \leqslant \tfrac{2}{3}\pi, \tag{8.56}$$

$$\mathrm{Bi}(\zeta) \sim \tfrac{1}{2}\pi^{-\frac{1}{2}}\zeta^{-\frac{1}{4}}\{\mathrm{i}\exp(-\tfrac{2}{3}\zeta^{\frac{3}{2}}) + \exp(\tfrac{2}{3}\zeta^{\frac{3}{2}})\} \text{ for } \tfrac{2}{3}\pi \leqslant \arg\zeta \leqslant \tfrac{4}{3}\pi, \tag{8.57}$$

$$\mathrm{Bi}(\zeta) \sim \tfrac{1}{2}\pi^{-\frac{1}{2}}\zeta^{-\frac{1}{4}}\{2\mathrm{i}\exp(-\tfrac{2}{3}\zeta^{\frac{3}{2}}) + \exp(\tfrac{2}{3}\zeta^{\frac{3}{2}})\} \text{ for } \tfrac{4}{3}\pi \leqslant \arg\zeta \leqslant 2\pi, \tag{8.58}$$

where a fractional power of ζ has the meaning given in the preceding section. An alternative form of (8.58) is

$$\mathrm{Bi}(\zeta) \sim \tfrac{1}{2}\pi^{-\frac{1}{2}}\zeta^{-\frac{1}{4}}\{-\mathrm{i}\exp(-\tfrac{2}{3}\zeta^{\frac{3}{2}}) + 2\exp(\tfrac{2}{3}\zeta^{\frac{3}{2}})\} \text{ for } -\tfrac{2}{3}\pi \leqslant \arg\zeta \leqslant 0. \tag{8.59}$$

In these formulae the ranges of $\arg\zeta$ end on Stokes lines, as explained in the preceding section. There are both dominant and subdominant terms in all three ranges of $\arg\zeta$ and the asymptotic behaviour of Bi(ζ) is therefore considerably more complicated than for Ai(ζ). This is illustrated by the Stokes diagram for Bi(ζ) given in fig. 8.6(b).

It is easily shown that the Stokes equation (8.7) is unaltered when ζ is replaced by $\zeta\exp(\frac{2}{3}\mathrm{i}\pi)$ or $\zeta\exp(\frac{4}{3}\mathrm{i}\pi)$. Hence $\mathrm{Ai}(\zeta\mathrm{e}^{\frac{2}{3}\mathrm{i}\pi})$ and $\mathrm{Ai}(\zeta\mathrm{e}^{\frac{4}{3}\mathrm{i}\pi})$ are solutions of the Stokes equation. For some purposes it is convenient to use $\mathrm{Ai}(\zeta\mathrm{e}^{\frac{2}{3}\mathrm{i}\pi})$ as the second fundamental solution, instead of Bi(ζ). It readily follows from (8.52), (8.53) that the asymptotic approximation for $\mathrm{Ai}(\zeta\mathrm{e}^{\frac{2}{3}\pi\mathrm{i}})$ is

$$\mathrm{Ai}(\zeta\mathrm{e}^{\frac{2}{3}\pi\mathrm{i}}) \sim \tfrac{1}{2}\pi^{-\frac{1}{2}}\zeta^{-\frac{1}{4}}\mathrm{e}^{-\frac{1}{6}\pi\mathrm{i}}\exp(\tfrac{2}{3}\zeta^{\frac{3}{2}}) \text{ for } -\tfrac{4}{3}\pi \leqslant \arg\zeta \leqslant 0, \tag{8.60}$$

$$\mathrm{Ai}(\zeta\mathrm{e}^{\frac{2}{3}\pi\mathrm{i}}) \sim \tfrac{1}{2}\pi^{-\frac{1}{2}}\zeta^{-\frac{1}{4}}\mathrm{e}^{-\frac{1}{6}\pi\mathrm{i}}\{\exp(\tfrac{2}{3}\zeta^{\frac{3}{2}}) + \mathrm{i}\exp(-\tfrac{2}{3}\zeta^{\frac{3}{2}})\} \text{ for } 0 \leqslant \arg\zeta \leqslant \tfrac{2}{3}\pi. \tag{8.61}$$

The Stokes diagram for $\mathrm{Ai}(\zeta\mathrm{e}^{\frac{4}{3}\pi\mathrm{i}})$ is shown in fig. 8.6(c) and is obtained from that for Ai(ζ) (fig. 8.6(a)) by rotation through 120° clockwise.

It must be possible to express $\mathrm{Ai}(\zeta\mathrm{e}^{\frac{2}{3}\pi\mathrm{i}})$ and $\mathrm{Ai}(\zeta\mathrm{e}^{\frac{4}{3}\pi\mathrm{i}})$ as linear combinations of Ai(ζ) and Bi(ζ). It can be shown (see, for example, Miller, 1946), that

$$2\mathrm{e}^{-\frac{1}{3}\pi\mathrm{i}}\mathrm{Ai}(\zeta\mathrm{e}^{\frac{2}{3}\pi\mathrm{i}}) = \mathrm{Ai}(\zeta) - \mathrm{i}\,\mathrm{Bi}(\zeta), \quad 2\mathrm{e}^{\frac{1}{3}\pi\mathrm{i}}\mathrm{Ai}(\zeta\mathrm{e}^{\frac{4}{3}\pi\mathrm{i}}) = \mathrm{Ai}(\zeta) + \mathrm{i}\,\mathrm{Bi}(\zeta) \tag{8.62}$$

and

$$\mathrm{Ai}(\zeta) + \mathrm{e}^{\frac{2}{3}\pi\mathrm{i}}\mathrm{Ai}(\zeta\mathrm{e}^{\frac{2}{3}\pi\mathrm{i}}) + \mathrm{e}^{\frac{4}{3}\pi\mathrm{i}}\mathrm{Ai}(\zeta\mathrm{e}^{\frac{4}{3}\pi\mathrm{i}}) = 0 \tag{8.63}$$

8.19. Connection formulae, or circuit relations

Equations (8.52) and (8.53) show that in the special case when ζ is real, the asymptotic approximations for Ai(ζ) are different according as ζ is positive or negative. A formula which gives one asymptotic approximation when the other is

known is sometimes called a 'connection formula' or 'circuit relation'. For example, the connection formula for $\mathrm{Ai}(\zeta)$ is

$$\zeta^{-\frac{1}{4}}\{\exp(-\tfrac{2}{3}\zeta^{\frac{3}{2}}) + \mathrm{i}\exp(\tfrac{3}{2}\zeta^{\frac{3}{2}})\} \leftrightarrow \zeta^{-\frac{1}{4}}\exp(-\tfrac{2}{3}\zeta^{\frac{3}{2}}) \tag{8.64}$$

(arg $\zeta = \pi$ on left, and 0 on right). The use of the double arrow was introduced by Jeffreys (1924). The terms on the left are for negative ζ and those on the right for positive ζ. Clearly, in order to give a connection formula it is necessary to know the value of the Stokes multiplier.

The connection formula for $\mathrm{Bi}(\zeta)$ is written slightly differently thus

$$\zeta^{-\frac{1}{4}}\{\mathrm{i}\exp(-\tfrac{2}{3}\zeta^{\frac{3}{2}}) + \exp(\tfrac{2}{3}\zeta^{\frac{3}{2}})\} \to 2\zeta^{-\frac{1}{4}}\exp(\tfrac{2}{3}\zeta^{\frac{3}{2}}). \tag{8.65}$$

Since the line arg $\zeta = 0$ is a Stokes line it does not matter whether the subdominant term is present or not. But the dominant term alone, for positive ζ, is insufficient to determine the asymptotic behaviour for negative ζ. On the other hand, the asymptotic behaviour for negative ζ determines completely the behaviour for positive ζ. Hence, following Langer (1934), we use only a single arrow in (8.65). For a discussion of connection formulae, see for example Heading (1970a, 1977a, 1979).

8.20. Stratified ionosphere. Uniform approximation

The earlier sections of this chapter started with the assumption that in the ionosphere the function $q^2(z)$, (8.2) is an exactly linear function of z and this led to the Stokes equation (8.7). This assumption may often be approximately true for a small range of z, but it is never exactly true in practice. It is important, therefore, to study solutions of the differential equation (7.6) when q^2 has an isolated zero as in fig. 8.1(b), but is not exactly linear. This can be done by using the method of § 7.9. We seek a fictitious 'comparison' medium with a new 'height' variable ζ, instead of z. The transformation is given by (7.49)–(7.52). The second term in the square brackets of (7.52) is $\frac{1}{2}\mathscr{D}(z;\zeta)$ where $\mathscr{D}$ is the Schwarzian derivative, given by (7.56). The new variable ζ is to be chosen so that $\mathscr{D}$ is zero when $q^2(z)$ is exactly linear, and small when q departs slightly from linearity. The derivative $z' = \mathrm{d}z/\mathrm{d}\zeta$, used in (7.56), must therefore exist and be bounded. This means that ζ is an analytic function of z. Thus there are two conditions. First, for the equation (7.52) to reduce to the Stokes equation, the first term in the square brackets must be given by

$$(kqz')^2 = -\zeta \tag{8.66}$$

which leads to

$$\pm\tfrac{2}{3}\mathrm{i}\zeta^{\frac{3}{2}} = k\int^z q\,\mathrm{d}z. \tag{8.67}$$

Second, since ζ is an analytic function of z, the lower limit of the integral must be at $z = z_0$ where $q = 0$. Hence

$$\zeta = \left(\frac{3}{2}\mathrm{i}k\int_{z_0}^{z} q\,\mathrm{d}z\right)^{\frac{2}{3}}. \tag{8.68}$$

If q^2 is given exactly by the linear function (8.2), then (8.68) is the same as (8.6). The sign and the fractional power in (8.68) are chosen so that ζ is positive when $\mathrm{Re}(z-z_0)$ is positive. Let $w=\zeta(k^2a)^{-\frac{1}{3}}$. If q^2 is given by (8.3), in which b gives the curvature of the $q^2(z)$ curve, then it can be shown that

$$\left.\begin{aligned} z-z_0 &= w+\frac{1}{5}\frac{b}{a}w^2+\frac{88}{700}\left(\frac{b}{a}\right)^2 w^3+\cdots, \\ w &= z-z_0-\frac{1}{5}\frac{b}{a}(z-z_0)^2-\frac{32}{700}\left(\frac{b}{a}\right)^2(z-z_0)^3-\cdots. \end{aligned}\right\} \tag{8.69}$$

Now (8.69) is substituted in (7.56) to give

$$\left.\begin{aligned} |\tfrac{1}{2}\mathscr{D}| &= \frac{1}{k^2}\left|\frac{3\zeta}{4q^4}\left(\frac{\mathrm{d}q}{\mathrm{d}z}\right)^2-\frac{\zeta}{2q^3}\frac{\mathrm{d}^2q}{\mathrm{d}z^2}+\frac{5k^2}{16\zeta^2}\right| \\ &= \frac{1}{k^2}\left|\frac{5\zeta}{16q^6}\left\{\frac{\mathrm{d}(q^2)}{\mathrm{d}z}\right\}^2-\frac{\zeta}{4q^4}\frac{\mathrm{d}^2(q^2)}{\mathrm{d}z^2}+\frac{5k^2}{16\zeta^2}\right| \end{aligned}\right\} \tag{8.70}$$

(compare (7.57)). Although the separate terms in (8.70) are infinite where $\zeta=0$, $q^2=0$, the combination of them is bounded. It can be shown that (8.70) is zero if $q^2(z)$ is exactly linear as in (8.2). If q^2 is given by (8.3) then it can be shown that

$$|\tfrac{1}{2}\mathscr{D}| = -\left(\frac{b}{a}\right)^2(k^2a)^{-\frac{2}{3}}\left\{\frac{9}{35}+\frac{809}{1050}\frac{b}{a}(z-z_0)+\cdots\right\}. \tag{8.71}$$

This confirms that when the curvature factor b of (8.3) is small, $|\frac{1}{2}\mathscr{D}|$ is very small, of order b^2.

If $\mathscr{D}$ is neglected, (7.52) is the same as the Stokes equation (8.7) with F for E. Its solution F must be some combination of $\mathrm{Ai}(\zeta)$ and $\mathrm{Bi}(\zeta)$ chosen to satisfy the physical conditions. When $z-z_0$ and thence ζ are large and positive $\mathrm{Ai}(\zeta)$ and $\mathrm{Bi}(\zeta)$ may be replaced by their asymptotic approximations. That for $\mathrm{Ai}(\zeta)$ is given by (8.52) and includes only the subdominant term $\zeta^{-\frac{1}{4}}\exp(-\frac{2}{3}\zeta^{\frac{3}{2}})$, which gets indefinitely smaller as ζ increases. That for $\mathrm{Bi}(\zeta)$ is given by (8.56) or (8.59) and includes the dominant term $\zeta^{-\frac{1}{4}}\exp(\frac{2}{3}\zeta^{\frac{3}{2}})$ which gets indefinitely larger as ζ increases. If the solution included any multiple of $\mathrm{Bi}(\zeta)$ the electric field would get larger and larger as the height z increased and the energy in the field would be extremely large at very great heights. This obviously could not happen in a field produced by radio waves incident on the ionosphere from below. Hence the solution cannot contain any multiple of $\mathrm{Bi}(\zeta)$ and so the physical conditions of the problem show that F is some multiple of $\mathrm{Ai}(\zeta)$. Further discussions of how to choose the correct solution are given in chs. 15, 18, 19.

The solution (7.51) is then

$$E \approx (z')^{\frac{1}{2}}\mathrm{Ai}(\zeta) \propto \zeta^{\frac{1}{4}}q^{-\frac{1}{2}}\mathrm{Ai}(\zeta) \tag{8.72}$$

where (8.66) has been used. For a level far enough below $z=z_0$, where ζ is negative, the asymptotic form (8.53) for $\mathrm{Ai}(\zeta)$ can be used, provided that $|\zeta| \gtrsim 1$ (compare

(8.29)). Now ζ is given by (8.68). Hence

$$E \approx q^{-\frac{1}{2}} \left\{ \exp\left(-ik \int_{z_0}^{z} q \, dz \right) + i \exp\left(ik \int_{z_0}^{z} q \, dz \right) \right\} \tag{8.73}$$

where a constant multiplying factor has been omitted. Thus the two terms in (8.73) are just the two W.K.B. solutions (7.26). The approximate solution (8.72) can be used near the level $z = z_0$ where the W.K.B. solutions fail, but it goes over continuously into the two W.K.B. solutions in a region where its asymptotic approximations can be used. It is an example of a 'uniform approximation'. It was first given, for the case of an isolated zero of q^2, by Langer (1937). Other examples of uniform approximations are mentioned in §§ 16.3, 17.3, 17.4.

The non-linearity of q^2 near its zero is important principally in the exponential terms of (8.73). These are the rapidly varying terms of the W.K.B. solutions, especially when k is large. If $q^2(z)$ is exactly linear, z' is independent of height z and therefore of ζ. When $q^2(z)$ is slightly non-linear as in the example of (8.3), where b is small, it can be shown from (8.69) that

$$|z'| = (k^2/a)^{\frac{1}{3}} \{1 + \tfrac{2}{5} b (a^2 k)^{-\frac{2}{3}} \zeta + \ldots.\}. \tag{8.74}$$

The use of (8.66) in (7.52), and thence of (8.72), rests on the assumption that z' does not change appreciably for values of $|\zeta|$ up to about unity. This requires that

$$|b(a^2 k)^{-\frac{2}{3}}| \ll 1. \tag{8.75}$$

For larger values of $|\zeta|$ the W.K.B. solutions can be used. Another way of deriving (8.75) was given by Budden (1961a, § 16.8).

The solution (8.72) is often needed in a form where the upgoing component wave, that is the first term in (8.73), has unit amplitude at the ground $z = 0$. This is achieved by using a constant multiplying factor with the second expression (8.72), thus

$$E = 2\zeta^{\frac{1}{4}} (\pi C/q)^{\frac{1}{2}} \mathrm{Ai}(\zeta) \exp\left(-ik \int_{0}^{z_0} q \, dz \right). \tag{8.76}$$

The results of this section and particularly (8.73) have now supplied the justification for the argument in § 7.19 that led to the expression (7.152) for the reflection coefficient, including the factor i which comes from the second term of (8.73). This means that the phase of the wave is advanced by $\frac{1}{2}\pi$ during the reflection process. An alternative explanation of this phase advance is given in the following section.

8.21. The phase integral method for reflection

In the formula (7.152) for the reflection coefficient, the integral is a contour integral along a path in the complex z plane, beginning at the ground $z = 0$ and ending at the reflection point $z = z_0$. It may be written in other forms. The function $q(z)$ is a two-valued function (7.2) with a branch point at z_0, and the two values will now be

written $q_1(z)$, $q_2(z)$. The value used in (7.152) is $q_1(z)$, chosen so that its real part is positive on the real z axis. It is therefore associated with the upgoing wave (7.149). The other value is $q_2(z) = -q_1(z)$, and if inserted in the exponential of (7.149) it would give the downgoing wave (7.150). Now (7.152) may be written

$$R = \mathrm{i}\exp\left\{-\mathrm{i}k\int_0^{z_0} q_1(z)\,\mathrm{d}z - \mathrm{i}k\int_{z_0}^{0} q_2(z)\,\mathrm{d}z\right\}. \tag{8.77}$$

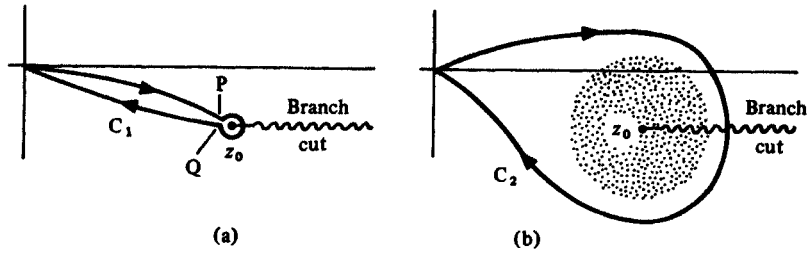
The contours for these two integrals need not be the same. They can be moved in different ways provided that no singularities are crossed.

Now let a branch cut be inserted in the complex z plane, running from z_0 to infinity as shown in fig. 8.8. Its exact position is not important. Then on crossing the cut q_1 and q_2 both change sign. The value of q_1 at a point on one side of the cut is the same as the value of q_2 at the adjacent point on the other side and vice versa. At the branch point z_0 both q_1 and q_2 are zero. The two contours in (8.77) can now be replaced by one contour C_1, fig. 8.8(a). The first part of C_1 runs from $z = 0$ to a point P very close to z_0 and the integrand is q_1. The next part is a very small circle that encircles the point z_0 clockwise from P to Q. Before the cut is crossed the integrand is q_1 and afterwards it is q_2. Since both q_1 and q_2 are small, the value of the integral on this circle can be made as small as desired by choosing the radius to be small enough. The third part of C_1 runs from Q back to $z = 0$ and the integrand is q_2. Then (8.77) becomes

$$R = \mathrm{i}\exp\left(-\mathrm{i}k\int_{\mathrm{C}} q\,\mathrm{d}z\right) \tag{8.78}$$

where C is C_1, and q, now written without a subscript, means q_1 before the cut is crossed and q_2 afterwards. The integrand in (8.78) is a continuous analytic function of z at all points on the contour. The contour C can therefore be distorted so that it is not very near the point z_0, as shown at C_2, fig. 8.8(b), provided that no singularities of q are crossed. The values of the integral and of R are unaffected. In particular it can be moved out into a region where the W.K.B. solutions are good approximations at all points on it, provided that this is not prevented by singularities of q that are too

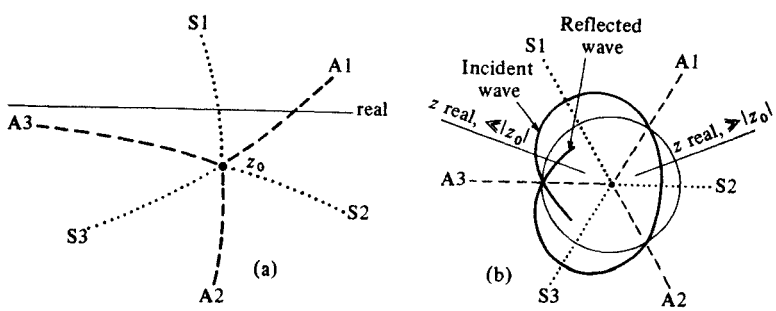
Fig. 8.8. The complex z plane showing possible contours C_1, C_2 for use with the phase integral formula (8.78). The shaded region in (b) is where $|\zeta| \lesssim 1$.



near to z_0. In the example of fig. 8.8(b) the region shaded with dots is where $|\zeta| \lesssim 1$. Outside this region the W.K.B. approximations can be used and this applies for all points on C_2. The integral now expresses the cumulative change of phase in the phase memory term of one W.K.B. solution as we proceed from the transmitter at the ground, $z = 0$, along the contour C_2 in the complex z plane and back to a receiver at $z = 0$. The integral is the phase integral and (8.78) is called the phase integral formula for reflection.

This process can be expressed in another way. The Stokes diagram for the function $\mathrm{Ai}(\zeta)$ in the uniform approximation solution (8.72) is shown in fig. 8.6(a). The Stokes lines $\arg\zeta = 0, \pm\frac{2}{3}\pi$, and the anti-Stokes lines $\arg\zeta = \pm\frac{1}{3}\pi, \pi$, radiate from the point $z = z_0$. They can be drawn in a diagram of the complex z plane as in fig. 8.9(a). If $q^2(z)$ were exactly linear they would be straight but in general they are curved. The Stokes diagram fig. 8.6(a) is shown again in fig. 8.9(b). Now consider the upgoing wave (7.149) near $z = 0$. It is one of the asymptotic forms of the uniform approximation solution (8.72). Its modulus gets larger as z gets more negative and therefore it is the dominant term. It is indicated by the label 'incident wave' in fig. 8.9(b). Now let the upper limit z of the integral in (7.149) move clockwise along the contour C_2 of fig. 8.8(b). The wave is at first the dominant term, but becomes subdominant after the anti-Stokes line A1 is crossed. The Stokes diagram fig. 8.9(b) shows that here there is no dominant term and for this wave the Stokes phenomenon does not occur near the Stokes line S2. After crossing the anti-Stokes line A2, the wave is again dominant. It becomes subdominant again after A3 is crossed, just before the real z axis is reached, and before the Stokes line S1 is crossed. Now it has been converted to the downgoing wave, indicated as 'reflected wave' in fig. 8.9(b). On the whole path the expression is a continuous analytic function of z that does not undergo the discontinuity associated with the Stokes phenomenon. This explains why the contour, in this example, must encircle the branch point in the clockwise sense. The upgoing and downgoing waves near the ground are thus linked by this continuous process. In some other applications of this method the branch point

Fig. 8.9. (a) is the complex z plane and shows the Stokes lines S and anti-Stokes lines A for the function $\mathrm{Ai}(\zeta)$. (b) is the Stokes diagram for this function.



must be encircled anticlockwise. In each problem of this kind the correct sense must be found by examining the relevant Stokes diagram.

Finally consider the factor $q^{-\frac{1}{2}}$ in (7.149). Now q^2 is analytic and may be represented by a series whose first two terms are (8.3). For the clockwise contours $\arg(q^2)$ decreases by 2π and so $\arg(q^{-\frac{1}{2}})$ increases by $\frac{1}{2}\pi$. Thus on returning to the initial point, the final value $q_2^{-\frac{1}{2}}$ is i times the initial value $q_1^{-\frac{1}{2}}$. Now in the expression (7.150) for the downgoing wave q means q_1 and $q^{-\frac{1}{2}}$ is therefore $-\mathrm{i}q_2^{-\frac{1}{2}}$. Thus the reflection coefficient R as given by (7.151) must be multiplied by i. This factor i is included in (7.152) and thence in (8.77), (8.78). It was used there because it appears in (8.73) as the Stokes multiplier of the Stokes equation. It has now been shown that it occurs as a direct consequence of the continuity of the W.K.B. solutions for the incident and reflected waves when the reflection branch point is encircled.

When the earth's magnetic field is allowed for, q_1 and q_2 are solutions of the Booker quartic, ch. 6, and in general $q_1 \neq -q_2$. The reflection point z_0 is where $q_1 = q_2$ but in general they are not zero here. For this case there is a uniform approximation solution (16.29)–(16.32), and the phase integral formula (8.78) can still be used. It can be applied to either the ordinary or the extraordinary wave. For a discussion of this see §§ 16.3, 16.4.

The validity of the phase integral formula for the reflection coefficient rests on the use of the function $\mathrm{Ai}(\zeta)$ in the uniform approximation solution (8.72) because the Stokes diagram, fig. 8.9(b), for this function contains a single continuous curve. This in turn implies that at real heights above the reflection level z_0 there is no downgoing wave. This might not be true if $q(z)$ has another singularity that is too close to z_0. For example, the reflection level z_0 where q is zero may be just below the maximum of an ionospheric layer. Then there is another point z_{00} just above the maximum where q is again zero. These two points are both branch points of $q(z)$. If they are too close together there may be no region between them where the W.K.B. solutions are good approximations, so that it cannot be asserted that there is no downgoing wave in the region above z_0 but below z_{00}. It is not then possible to find a contour C_2 as in fig. 8.8(b). This situation occurs when the electron concentration $N(z)$ has a parabolic distribution and the problem is discussed in §§ 15.9–15.11. Another example occurs for vertical polarisation and near vertical incidence, when q must be replaced by an 'effective' value Q, (15.27) that has a pole just above z_0. This case is discussed in §§ 15.5–15.7.

The first to use phase integral methods for radio propagation problems was T.L. Eckersley. His first main paper on the subject (Eckersley, 1931) is entitled 'On the connection between ray theory of electric waves and dynamics'. In the preceding years the analogy between the motion of a particle in a potential field, and the motion of a wave packet in a medium whose refractive index varies in space, had been used to develop the science of wave mechanics. It was postulated that a moving

particle is represented by a progressive wave whose phase must be a single valued function of the space coordinates. Thus when it goes round a closed orbit, the total change of phase, that is the 'phase integral', must be an integer times 2π. This is the origin of the term 'phase integral method'. Ray theory is closely analogous to classical mechanics with the addition of the concept of phase, leading to some of the quantum conditions.

Eckersley's first group of papers (1931, 1932a, b) on the subject was concerned with guided waves. The simplest kind of wave guide has two plane reflecting boundaries with free space between them. A single wave guide mode can be considered as two plane progressive waves in this space, with their wave normals making angles $\theta, \pi - \theta$ with the normal to the boundary planes (Brillouin, 1936; Chu and Barrow, 1938). For a self-consistent mode, the twice reflected wave must be identical with the original wave. This requires that the total change of complex phase in the double traverse of the guide width and at the two reflections must be an integer times 2π. It leads to the well known 'mode condition', and is a very simple example of the use of the phase integral. It is analogous to the phase integral condition for the wave associated with a particle in a closed orbit.

This simple example is exceptional because the reflections at the sharp boundary planes cannot be dealt with by ray theory. But in some other problems considered by Eckersley the reflections occurred in continuous slowly varying media, such as the ionosphere, where it was possible to solve the whole problem by ray theory. In this way the phase integral method developed into a method of calculating the reflection coefficient of a continuously varying stratified medium, as described earlier in this section, even when no mode condition is imposed.

For a review of the subject see Budden (1975). For discussion of the mathematical details see Heading (1962a, c, 1976, 1977b), Wait (1962), Evgrafov and Fedoryuk (1966).

The W.K.B. solutions, and the phase integral formula for reflection (7.152) or equivalently (8.78) are the basis of ray tracing methods, chs. 10, 12–14, and are normally used for studying propagation and reflection of high frequency waves in the ionosphere and magnetosphere. In practice this means frequencies greater than about 100 to 200 kHz, or possibly less in the upper ionosphere and magnetosphere when the medium changes little within one free-space wavelength. For high frequencies the methods of ray theory are the most convenient, and the only sensible way of studying propagation. It is shown in §§ 10.17–10.19 that a ray can be traced right through a region near where it touches a caustic surface, and this is effectively a region of reflection. The use of ray theory in this way is possible when the reflection points are well isolated, that is well separated in complex space from others. The only additional feature of the phase integral formula is the factor i in (7.152), (8.78) but this is not important in ray theory. Coupling points, like reflection points, are

points in the complex z plane where two roots of the Booker quartic are equal. It is shown in § 16.7 that coupling between two characteristic waves, ordinary and extraordinary, can be treated by the phase integral method, and if the coupling point is sufficiently isolated, the process is very similar to the phase integral method for reflection. Mathematically, the two processes are the same, and the term 'coupling point' is used to mean either a coupling or a reflection point.

In this sense, therefore, the phase integral method is a widely used physical principle that is essentially the same as ray theory. The main purpose of discussing it in this book is to see how far it can be extended and, most important of all, to see when it fails and why.

When two coupling points are close together but can be studied separately from others, this introduces ideas that are the basis of new physical principles going beyond simple ray theory; for example, partial penetration and reflection §§ 15.10, 17.3, radio windows §§ 17.6–17.9, limiting polarisation §§ 17.10, 17.11. But for computing it is now simpler and safer to use numerical integration of the basic differential equations; chs. 18, 19. The phase integral method uses transformations of the variables that introduce, at the coupling points, singularities that are not present in the original basic differential equations. These transformations are useful only as a means of helping to explain the results in terms of physical processes. They are best avoided in numerical integration of the differential equations.

When three or more coupling points need to be considered together, the phenomena are now too complicated to be interpretable in terms of simple physical processes. It is still possible to use the phase integral method in some special cases; see, for example, Heading (1977b). But a general rule that will cover all cases has not been formulated. For very low frequencies, therefore, and for other cases where ray theory is suspect because of the proximity of coupling points, it is safest and also easiest to use full wave numerical solutions.

8.22. The intensity of light near a caustic

The integral (8.19) was originally derived by Airy (1838, 1849) in a study of the variation of the intensity of light near a caustic. Caustics can also be formed by radio waves, see §§ 10.17–10.23, and it is therefore useful to give the theory. Huyghens's principle is used to construct 'amplitude–phase' diagrams for a series of neighbouring points. Each diagram is a spiral, somewhat analogous to Cornu's spiral, and its vector resultant gives the amplitude and phase of the light at the point considered.

Consider a parallel beam of light incident from the left on a convex lens, fig. 8.10. For simplicity the lens is assumed to be cylindrical, and the problem is treated as though it is in two dimensions only. It is well known that the rays of the beam are tangents to a caustic curve XY. It is required to find the intensity of the light at the points of a line AB perpendicular to the caustic. Let Q be a typical point of this line.

Consider a plane wave front of the beam (RS in fig. 8.10), just before it reaches the lens. Imagine this wave front to be divided into infinitesimal strips of equal width δs, which all radiate cylindrical waves of the same intensity. On arrival at Q these waves all have the same amplitude, but their phases depend on the position, s, of the strip from which each originates. If Q is inside the caustic, as shown in fig. 8.10, two geometrical rays pass through it. Assume that one of these, RQ, is included in the narrow pencil between rays 1 and 2, and the other, SQ, in the pencil between rays 4 and 5.

The pencil enclosed by the rays 1 and 2 comes to a focus at X. The point Q is within this pencil, and the light reaches it before it reaches the focus X. It is then easily shown that for the ray RQ within this pencil Fermat's principle is a principle of least time. The variation with s of the phase $\phi(s)$ therefore has a minimum for the Huyghens wavelets that originate near R. The pencil enclosed by the rays 4 and 5 comes to a focus at Y. The point Q is within the pencil and beyond the focus, and it can then similarly be shown that for the ray SQ Fermat's principle is a principle of greatest time, so that the phase $\phi(s)$ has a maximum for Huyghens wavelets that originate near S. The function $\phi(s)$ has no other turning points (fig. 8.11, curve Q). If Q is outside the caustic, at T say, no geometrical rays pass through it. Then $\phi(s)$ has no turning points, and varies as shown in fig. 8.11, curve T. If Q is on the caustic, at W say, the two turning points coincide (fig. 8.11, curve W).

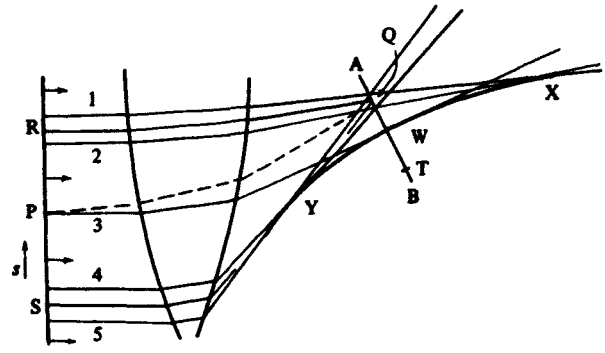
Let the zero of s be chosen midway between the two turning points, or at the point of least slope when there are no turning points. Then the form of the curves in fig. 8.11 is given by the cubic law

$$\phi = K(\zeta s + \tfrac{1}{3}s^3), \tag{8.79}$$

where ζ depends on the distance of Q from the caustic, and K is a constant. Clearly the scale of s can be chosen so that $K = 1$. When Q is on the illuminated side of the caustic, as in fig. 8.10, ζ is negative, and when Q is outside the caustic ζ is positive.

The value of ζ can be found as follows. Let the distance QW $= u$, and let the radius

Fig. 8.10. Caustic formed by a beam of light incident on a lens.



of curvature of the caustic be R. For curve 1, in fig. 8.11, ζ is negative, and the difference between the maximum and minimum values of ϕ is $\frac{4}{3}\zeta^{\frac{3}{2}}$ (with $K = 1$). This is the difference between the phases of the light arriving at Q via the ray pencils 4–5 and 1–2. Now any two rays that are close together, such as 4 and 5, have the same phase where they meet on the caustic. Hence the phase difference between the waves arriving at X and Y is just the phase difference that would arise from the curved path YWX along the caustic, namely $4\pi R\theta/\lambda$ where 2θ is the angle subtended by the arc XY at its centre of curvature. Thus the difference between the phases of the light arriving at Q via the two ray pencils is

$$\frac{2\pi}{\lambda}\{\mathrm{YQ} + \mathrm{QX} - \mathrm{arc(YWX)}\} = 4\pi R(\tan\theta - \theta)/\lambda \approx \frac{4\pi R\theta^3}{3\lambda}, \tag{8.80}$$

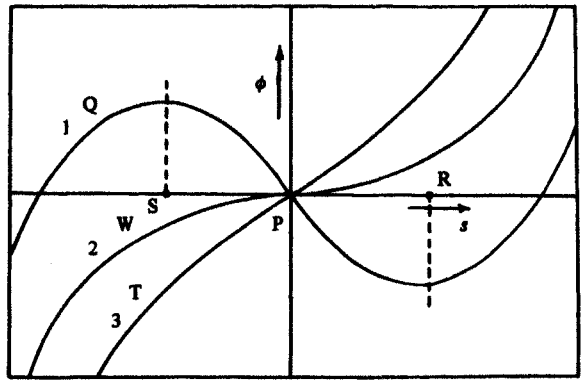
so that $\zeta^3 \approx \pi^2 R^2\theta^6/\lambda^2$. Now $\mathrm{QW} = u \approx \frac{1}{2}R\theta^2$, so that when u is small

$$\zeta = 2u\left(\frac{\pi^2}{R\lambda^2}\right)^{\frac{1}{3}}. \tag{8.81}$$

Clearly ζ could be expressed by an ascending power series in $u = \mathrm{QW}$, and if ζ is small enough, powers higher than the first can be neglected.

For each position of the point Q an amplitude–phase diagram may now be constructed. The curvature of this at a given point is proportional to the slope at the corresponding point in the ϕ–s curve. Since this slope eventually increases indefinitely both for s increasing and for s decreasing, the amplitude–phase diagrams end in spirals. Typical diagrams are shown in fig. 8.12, corresponding to the three curves of fig. 8.11. In each case the closing vector of the diagram is marked with an arrow. It gives the resultant amplitude of the light that reaches Q. There is an obvious analogy between these diagrams and Cornu's spiral, for which the ϕ–s curve is a parabola. Cornu's spiral always has the same form. In diffraction problems whose solution involves a Cornu spiral, variations of intensity occur because

Fig. 8.11. Dependence of phase $\phi(s)$ of Huyghens wavelet on position s of source on initial wavefront.

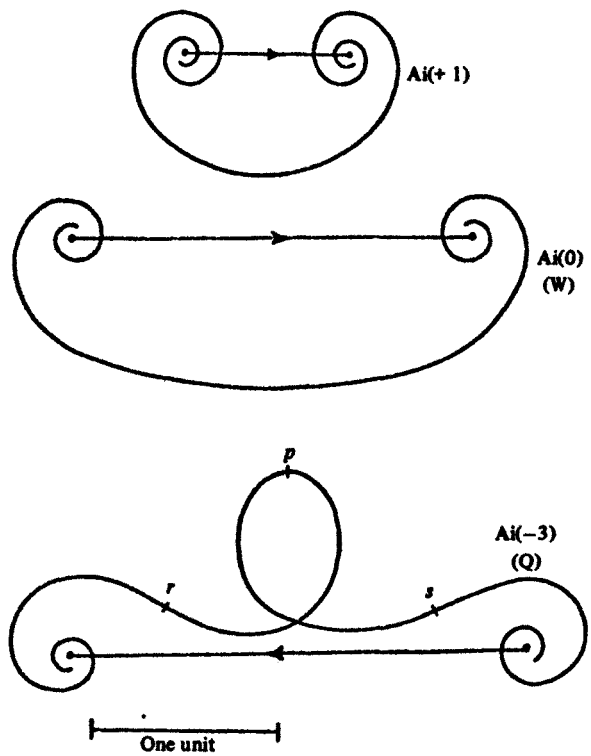


different fractions of the spiral are used to determine the intensity reaching different points. In the present problem however, the shape of the spiral is different for different values of ζ. For all points of the line AB, the resultant of the whole of the spiral is found, but the intensity varies because the shape of the spiral varies from point to point.

In fig. 8.12, let the x axis be chosen parallel to the element corresponding to $s = 0$. Then the diagrams are symmetrical about the y axis, and the resultant vector is always parallel to the x axis. When ζ is positive (first diagram of fig. 8.12), the curvature of the spiral always has the same sign and the length of the resultant decreases monotonically as ζ increases. When ζ is negative (third diagram of fig. 8.12), the middle section of the spiral has opposite curvature from the rest, and as ζ decreases the limiting points of the spiral repeatedly move completely round the origin. The resultant thus alternates in sign, and in the third diagram of fig. 8.12 it is negative.

The contribution of an element δs of the wave front to the resultant vector is proportional to $\delta s \cos \phi$, and hence the resultant vector is proportional to

Fig. 8.12. Spirals associated with the Airy integral function. In the third spiral the points p, r, s, correspond roughly to the points P, R, S in figs. 8.10 and 8.11.



$$\int_{-\infty}^{\infty} \cos\phi \, ds = \int_{-\infty}^{\infty} \cos(\zeta s + \tfrac{1}{3}s^3)\, ds. \tag{8.82}$$

Apart from a constant, this is the standard expression (8.19) for the Airy integral function, Ai(ζ).

When Q is well beyond the caustic on the illuminated side (ζ large and negative), the turning points of the ϕ–s curve are widely separated, and near each turning point the curve is approximately a parabola. Hence the portion of the amplitude–phase diagram corresponding to the neighbourhood of each turning point is approximately a Cornu spiral. The two Cornu spirals curve opposite ways, and together constitute the more complex spiral whose resultant is the Airy integral function; see fig. 8.13.

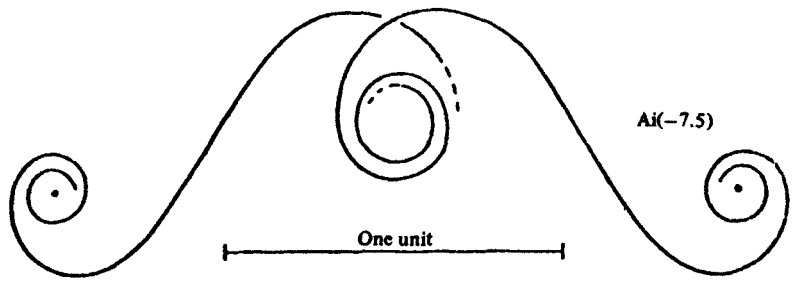
PROBLEMS 8

8.1. A linear second order differential equation with independent variable z has two solutions $S_1(z)$, $S_2(z)$. A prime $'$ is used to denote d/dz. The function $S_1S_2' - S_2S_1' = W(S_1, S_2)$ is called the Wronskian for these solutions. Prove that if the differential equation has no first derivative term, any Wronskian is a constant.

For the Stokes differential equation $\mathrm{d}^2E/\mathrm{d}z^2 = zE$ prove that $W(\mathrm{Ai}, \mathrm{Bi}) = 1/\pi$. If S_1, S_2 are the two series solutions in (8.8) find (a) $W(S_1, S_2)$, (b) $W(S_1, \mathrm{Ai})$, (c) $W(S_2, \mathrm{Ai})$. (Express (b) and (c) in terms of Ai(0), Ai$'$(0)).

8.2. The functions Ai(z) and Bi(z) (fig. 8.5) are both real and oscillatory when z is real and negative. They can be thought of as 'standing waves', analogous in some ways to $\cos kz$, $\sin kz$. It is sometimes more convenient to use solutions which are analogues of progressive waves. This suggests the use of $\mathrm{C}_+ = \mathrm{Ai} + \mathrm{i}\,\mathrm{Bi}$, $\mathrm{C}_- = \mathrm{Ai} - \mathrm{i}\,\mathrm{Bi}$. Prove that $\mathrm{C}_+(z) = -2\Omega^2\,\mathrm{Ai}(\Omega^2 z)$, and $\mathrm{C}_-(z) = -2\Omega\,\mathrm{Ai}(\Omega z)$ where $\Omega = \exp(\frac{2}{3}\pi\mathrm{i})$. (Use the contour integrals (8.16), (8.17)). Show that for real negative z their asymptotic forms are proportional to $z^{-\frac{1}{4}}\exp(\mp\frac{2}{3}\mathrm{i}|z|^{\frac{3}{2}})$ respectively (upper sign for C_+). Hence confirm that they represent progressive waves.

Fig. 8.13. Spiral associated with Airy integral function of large negative argument, showing resemblance to two Cornu spirals. The part of the left half where overlapping occurs is not shown.



8.3. Find solutions $S(z)$ of the Stokes equation which, if possible, satisfy the following conditions. Express $S(z)$ in terms of $\mathrm{Ai}(z)$ and $\mathrm{Bi}(z)$, whose Wronskian is $1/\pi$. When there is no possible solution explain why. (a is a number that makes $\mathrm{Ai}(-a)=0$.)

(a) $S(0)=0, \quad S'(0)=-1.$

(b) $S(0)=1, \quad S'(0)=0.$

(c) $S(-1)=0, \quad S(0)=1.$

(d) $S(0)=0, \quad S(1)=0.$

(e) $S(0)=1, \quad S(\infty)=0.$

(f) $S(0)=0, \quad S(\infty)=0.$

(g) $S(-a)=0, \quad S(\infty)=0.$

(h) $S(-a)=0, \; S'(-a)=1.$

(i) $S(-a)=1, \quad S(\infty)=0.$

8.4. The function $y(z)$ satisfies $\mathrm{d}^2y/\mathrm{d}z^2=(z-a)y$ where a is an adjustable parameter, constant for any one solution. Find those values of a for which solutions are possible in the following cases, and express the solutions in terms of Ai and Bi. In each case state whether the given boundary conditions are homogeneous or inhomogeneous.

(a) $y(0)=0$, $y(\infty)$ bounded. y has no zeros for $z>0$.

(b) $y(0)=0, \; y(1)=0.$

(c) $y(0)=1$, $y(\infty)$ bounded.

(d) $y'(0)=0$, $y(\infty)$ bounded.

(e) $y'(0)=0, \; y'(1)=0.$

(f) $y'(0)=1, \; y(1)=0.$

8.5. Find a second order differential equation satisfied by $u=\mathrm{d}y/\mathrm{d}z$ where y can be any solution of the Stokes equation.

8.6. The differential equation $\mathrm{d}^2y/\mathrm{d}z^2=\frac{1}{4}z^2y$ is a special case of Weber's equation. Find (a) the W.K.B. solutions, (b) the Stokes lines, (c) the anti-Stokes lines and (d) the Stokes multipliers.