Appendix III

Asymptotic approximations to integrals

THE purpose of this appendix is to provide the mathematical background to methods of some generality, referred to in the main text, for obtaining asymptotic approximations to certain types of integral that frequently occur in optics.

1 The method of steepest descent

This is a method* for obtaining, for large values of k, asymptotic approximations to complex integrals of the type

$$\int g(z)e^{kf(z)} dz, \tag{1}$$

where g(z) and f(z) are independent of k.

For present purposes k is to be considered real and positive, and is assumed so in the later discussion. In fact, however, the results to be given are in general also true for complex values of k, and it is desirable to make a few prefatory remarks about asymptotic expansions of a function of a complex variable, ζ , say.

First, a definition of what is meant by an asymptotic expansion, due to Poincaré,† may be stated as follows: if

$$F(\zeta) = \sum_{m=0}^{n} \frac{a_m}{\zeta^m} + R_n(\zeta), \tag{2}$$

where, for all n, $\zeta^n R_n(\zeta) \to 0$ as $\zeta \to \infty$ for arg ζ within a given interval, a_0 , a_1, \ldots, a_n being constants, then one writes

$$F(\zeta) \sim a_0 + \frac{a_1}{\zeta} + \frac{a_2}{\zeta^2} + \cdots,$$
 (3)

and the right-hand side of (3) is called an asymptotic expansion of $F(\zeta)$ for the given range of arg ζ .

^{*} Essentially due to P. Debye, Math. Ann., 67 (1909), 535. † H. Poincaré, Acta Math., 8 (1886), 295.

If $F(\zeta)$ is the quotient of two functions, say $G(\zeta)$ and $H(\zeta)$, one can write

$$G(\zeta) \sim H(\zeta) \left(a_0 + \frac{a_1}{\zeta} + \frac{a_2}{\zeta^2} + \cdots \right).$$
 (4)

In fact it should be stressed that further on in the discussion the expansions considered are in the form of the right-hand side of (4), $H(\zeta)$ being $\exp(a\zeta)$, where a is some constant.

Some of the main properties of asymptotic expansions are now briefly set out.* If the series in (3) terminates, or converges for sufficiently large $|\zeta|$, it is asymptotic; but it often fails to converge for any value of $|\zeta|$. In general, for a given $F(\zeta)$ a particular expansion holds only for a specified range of arg ζ ; if it holds for all arg ζ then it converges. Again, for a given $F(\zeta)$ in the appropriate range of arg ζ the asymptotic expansion is unique, in the sense that the coefficients in (3) are unique; on the other hand the asymptotic expansion of any one function also belongs to an infinity of other functions, that of $F(\zeta) + e^{-\zeta}$, for example, being the same as that of $F(\zeta)$ for $-\frac{1}{2}\pi < \arg \zeta < \frac{1}{2}\pi$. The asymptotic expansion of the product of two functions is obtained by multiplying out their respective asymptotic expansions. Finally (3) may be integrated term by term to yield unconditionally the asymptotic expansion of the integral of $F(\zeta)$; it also may be differentiated term by term to yield, if it exists, the asymptotic expansion of the differential of $F(\zeta)$.

In general, for a given (sufficiently large) value of $|\zeta|$, the moduli of the terms of (3) start by decreasing successively to a minimum and then subsequently increase. Roughly speaking, if the expansion is summed to any term prior to the smallest, the error is of the order of the first omitted term.† Evidently, the larger $|\zeta|$ is, the greater is the available accuracy. In physical applications it is often sufficient to use the first term only; to borrow an illustration from electromagnetic theory, the radiation field of a finite source distribution is the first term of the asymptotic expansion of the complete field in inverse powers of distance from the source.

The basis of the method for obtaining the asymptotic expansion of (1) in inverse powers of k is to relate it to integrals of the form.

$$\int_0^\infty h(\mu) e^{-k\mu^2} d\mu. \tag{5}$$

The asymptotic development of (5) can be easily derived; it results from expanding $h(\mu)$ as an ascending power series in μ and carrying out the integration term by term.

This last remark is, in fact, the substance of *Watson's lemma*,§ which may be stated in the following form: Let

^{*} See, for example, H. and B. S. Jeffreys, *Methods of Mathematical Physics* (Cambridge, Cambridge University Press, 1946), Chapter 17; or A. Erdélyi, *Asymptotic Expansions* (New York, Dover Publications, 1956).

[†] For a method of estimating the error see H. and B. S. Jeffreys, loc. cit., Chapter 17.

The choice of μ^2 in the exponential of the integrand rather than the first (or some other) power of μ is of no great significance; an even power is convenient because it oftens happens that the path of integration in (1) begins and ends at infinity in such a way that (1) can be transformed to (5) with the lower limit replaced by $-\infty$.

[§] G. N. Watson, *Theory of Bessel Functions* (Cambridge, Cambridge University Press, 2nd edition, 1948), Chapter 8. See also H. and B. S. Jeffreys, *loc. cit.*, Chapter 17.

$$h(\mu) = \frac{1}{\mu^{\alpha}} \sum_{s=0}^{\infty} c_s \mu^{\beta s},\tag{6}$$

with radius of convergence ρ , where β is real and positive and the real part of α is less than 1. Let a real number d exist such that $\mu^{\alpha} e^{-d\mu^{2}} h(\mu)$ is bounded for all real values of μ greater than ρ . Then (with Γ denoting the gamma function)

$$\frac{1}{2k^{\frac{1}{2}(1-\alpha)}} \left[c_0 \Gamma\left(\frac{-\alpha+1}{2}\right) + c_1 \Gamma\left(\frac{\beta-\alpha+1}{2}\right) \frac{1}{k^{\frac{1}{2}\beta}} + c_2 \Gamma\left(\frac{2\beta-\alpha+1}{2}\right) \frac{1}{k^{\beta}} + c_3 \Gamma\left(\frac{3\beta-\alpha+1}{2}\right) \frac{1}{k^{\frac{3}{2}\beta}} + \cdots \right]$$
(7)

is the asymptotic expansion of (5).

In cases of physical interest it is likely that α will be zero and β unity. Often too, as already suggested, the first term of (7) provides an adequate approximation.

In order to be able to change the variable of integration so that (1) is represented by one or more integrals of the type (5) it is necessary that the path of integration should be made up of sections along each of which the imaginary part of f(z) is constant and the real part of f(z) decreases monotonically to $-\infty$. If this is not the case, the first step is to distort the path appropriately. The path distortion is, of course, governed by the fundamental rules of integration in the complex plane; here it will merely be shown how it may be possible to choose a path, by means of sections having the required properties, with the supposition that the evaluation of any remaining contour integral is to be attempted by standard techniques.

Some quite general remarks can be made about paths along which the imaginary part of f(z) is constant, irrespective of the nature of the particular function. These will suffice to indicate how two points in the complex plane are joined in the desired manner, though the method may perhaps fail if f(z) has singularities. In the present connection an especially important role is played by the points at which

$$\frac{\mathrm{d}f}{\mathrm{d}z} = 0. \tag{8}$$

These are called *saddle-points*, because they are the points where the real and imaginary parts of f(z) are stationary with respect to position in the complex plane, without being absolute maxima or minima.*

Now if

$$f(z) = u(x, y) + iv(x, y)$$
(9)

it is easy to show from the Cauchy-Riemann relations that along any path v(x, y) = constant the rate of change of u(x, y) can only vanish at a saddle-point. In other words, on a path v(x, y) = constant which does not traverse a saddle-point, u(x, y) is strictly monotonic throughout and, on a path v(x, y) = constant which goes through one or more saddle-points, u(x, y) is strictly monotonic between adjacent saddle-points and from the terminal saddle-points to the respective ends of the path. In

^{*} The real and imaginary parts of an analytic function have no absolute maxima or minima in the complex plane.

fact, from any point, the direction in which u(x, y) decreases most rapidly is one along which v(x, y) is constant, and in this sense such paths are paths of *steepest descent*. Provided f(z) is single-valued, or is made so by the use of branch-cuts, it follows that, starting from any point (x_0, y_0) , a path $v(x, y) = v(x_0, y_0)$ can be chosen which continues always in the direction of decreasing u(x, y) and terminates either at infinity or at a singularity. It is readily seen that, apart from singularities, the only points at which a path v(x, y) = constant can go in more than one direction are where df/dz = 0; thus, once the path $v(x, y) = v(x_0, y_0)$ has been started correctly, with u(x, y) decreasing, no question of an alternative route arises unless a saddle-point is encountered, in which case at least one direction of decreasing u(x, y) will be available.

Now suppose that the end points of the given path of integration in (1) are A and B, and that it is possible to find paths of steepest descent from A and B, respectively, to infinity. If both these paths terminate at infinity in the same section of convergence of the integral, then the procedure is complete; but if the terminations at infinity are in different sections of convergence, then one must be joined to the other by a path v(x, y) = constant along which the rate of change of u(x, y) changes sign only once (at a saddle-point), or if necessary by several such paths via intermediary regions at infinity. An asymptotic expansion, in the extended sense of (4), can be obtained for each of the different paths involved, but the true asymptotic expansion of the original integral is simply that for the path on which u(x, y) attains its greatest value.* Similar remarks apply when the paths of steepest descent terminate in singularities.

It should be noted that there are cases outside the precise scope of the method just outlined which may nevertheless be treated by a slight modification. As a preliminary it is remarked that replacement of the upper limit in (5) by any positive number (independent of k) leaves the asymptotic expansion (7) unaltered. Thus the case can be treated involving a path of steepest descent along which u(x, y) tends down to a finite value instead of $-\infty$. Again, it is possible to use a path v(x, y) econstant traversing more than one point at which the rate of change of u(x, y) changes sign.

Perhaps the commonest situation where the method of steepest descent is applicable is that in which the path of integration v(x, y) = constant runs from a saddle-point to infinity with u(x, y) decreasing monotonically all the way.† A well-known formula for the first term of the asymptotic expansion is then readily derived. For suppose the saddle-point is at z_0 , and change the variable of integration in (1) by the transformation

$$f(z) = f(z_0) - \mu^2. (10)$$

Then (1) becomes

$$-2e^{kf(z_0)} \int_0^\infty \frac{g(z)}{f'(z)} \mu e^{-k\mu^2} d\mu.$$
 (11)

^{*} It is, of course, possible that the greatest value of u(x, y) will be attained on more than one of the paths.

[†] Strictly the most common case is that in which the path begins and ends at infinity with u(x, y) running monotonically from $-\infty$ up to a maximum at a saddle-point, and then monotonically from the maximum down to $-\infty$. Formula (12) holds in this case if multiplied by 2, but for the subsequent discussion it is more convenient to take a path starting at the saddle-point. Because the case in which it is arranged that the path of integration starts from or traverses a saddle-point is so common, the method of steepest descent is sometimes called the *saddle-point method*.

To obtain the first term in the asymptotic expansion of (11) the value of $\mu g(z)/f'(z)$ at $\mu=0$ ($z=z_0$) is required. Provided $f''(z_0)$ is not zero and $g(z_0)$ is not infinite this is easily seen to be $g(z_0)/\sqrt{-2f''(z_0)}$, where the sign of the square root must be determined by an examination of each particular case. The required approximation is therefore

$$\sqrt{-\frac{\pi}{2f''(z_0)}}g(z_0)\frac{e^{kf(z_0)}}{\sqrt{k}}.$$
 (12)

A few remarks should be added regarding the cases in which the approximation (12) needs supplementing or is invalid. The full asymptotic expansion is, of course, obtained by expanding $\mu g(z)/f'(z)$ as a power series in μ and carrying out the integration in (11) term by term. As indicated in (7), the power of k appearing in the first term of the asymptotic expansion is determined by the power of μ with which the series expansion of $\mu g(z)/f'(z)$ begins. If the first term of this series expansion is $A\mu^{-p}$, where p must be less than unity for the integral to converge, then the first term of the asymptotic expansion is

$$-A\Gamma\left(\frac{-p+1}{2}\right)\frac{e^{kf(z_0)}}{k^{\frac{1}{2}(1-p)}}. (13)$$

Thus if $g(z_0)$ is infinite or $f''(z_0)$ is zero, (12) is invalid and must be replaced by (13). In these cases the factor multiplying $\exp[kf(z_0)]$ tends to zero less rapidly than $k^{-1/2}$, as k tends to infinity. If $g(z_0) = 0$ or $f''(z_0)$ is infinite (12) is not incorrect, but then merely states that the factor multiplying $\exp\{kf(z_0)\}$ tends to zero more rapidly than $k^{-1/2}$, as k tends to infinity.

If the path of integration is one of steepest descent running to infinity as before but starting from a point which is not a saddle-point, it is easy to see that the non-exponential part of the first term of the asymptotic expansion is, in general, proportional to k^{-1} , in contrast to the factor $k^{-1/2}$ of (12). Again, though, if g(z)/f'(z) is singular or zero at the end point, the power of k is different, and depends on the order of the singularity or zero.

What has been said so far covers fairly comprehensively the results yielded by the method of steepest descent for asymptotic expansions in the strict mathematical sense, in which k is allowed to become indefinitely great, and other parameters are supposed to have prescribed values independent of k. But it has been shown that the form of the asymptotic approximation depends on particular conditions, for example, whether the path of steepest descent starts at a saddle-point or not. That is, it may depend on parameters other than k in the sense that it changes abruptly when these parameters take certain critical values.* Thus for any *given* value of k, no matter how large, if the values of other parameters are *sufficiently* close to those for which the asymptotic form would be different, the formulae mentioned above clearly fail to provide good numerical approximations. Hence there is a practical requirement for further expressions which will give smooth transitions from one asymptotic form to another. These are naturally more elaborate functions of k than (13), but it is worth mentioning three

^{*} This is essentially a matter of *nonuniform* convergence. To take a trivial illustration, as $k \to \infty$, $1/(1+ka) \sim 1/ka$ for $a \ne 0$, and ~ 1 for a = 0.

cases which can be satisfactorily treated. In the brief outline which follows it is tacitly assumed in each case that the only complexity is the specific one under discussion.

First, suppose that $f''(z_0)$ approaches zero. Then (12) is only a good approximation for increasingly large values of k, and an expression is required to effect the transition from (12) to the different form which is appropriate when $f''(z_0) = 0$. Since $f''(z_0)$ is nearly zero there must be a second saddle-point, z_1 say, near z_0 . Then the required expression can be obtained by making the transformation

$$f(z) = \alpha - \beta \mu + \frac{1}{3}\mu^3,$$

where

$$\alpha = \frac{1}{2}[f(z_1) + f(z_0)], \qquad \frac{2}{3}\beta^{\frac{3}{2}} = \frac{1}{2}[f(z_1) - f(z_0)].$$

This transformation gives z as a regular function of μ in the vicinity of z_0 and z_1 , and leads to an asymptotic development expressed in terms of the Airy integral and its first derivative with argument* $k\beta$.

Next, suppose that the expansion of $\mu g(z)/f'(z)$ as a power series in μ has a radius of convergence which approaches zero because g(z) has a simple pole which gets close to the saddle-point. Again, (12) is only a good approximation for increasingly large values of k, and an expression is required to effect the transition from (12) to the different form which is appropriate when the pole is at the saddle-point. This case has been treated by various authors.† The idea is to write the nonexponential part of the integrand as the sum of two terms, one containing the pole only, the other being regular. The latter can then be handled in the usual way and the former yields a Fresnel or error integral, with, in general, a complex argument.

Finally, suppose that the starting point of a path of steepest descent is not a saddle-point, but approaches very close to one. In this case it is clear that the error integral can again be used to link the two asymptotic forms.:

2 The method of stationary phase§

This is an alternative, if very similar, method to that of steepest descent. Though perhaps less general and less immediately convincing analytically, it often has the advantage of closer contact with the physical problem. The integral to be considered is more appropriately written

$$\int g(z)e^{ikf(z)} dz \tag{14}$$

^{*} The role of the Airy integral in uniform approximations of this type has long been recognized (see, for example, H. and B. S. Jeffreys, *loc. cit.*, Chapter 17), but only more recently has a satisfactory method, as outlined, been fully explored. (See C. Chester, B. Friedman and F. Ursell, *Proc. Camb. Phil. Soc.*, **53** (1957), 599.)

[†] W. Pauli, *Phys. Rev.*, **54** (1938), 925. H. Ott, *Ann. d. Physik* (5), **43** (1943), 393. P. C. Clemmow, *Quart. J. Mech. Appl. Maths.*, **3** (1950), 241; *Proc. Roy. Soc.*, A, **205** (1951), 286. B. L. van der Waerden, *Appl. Sci. Res., Hague*, B, **2** (1952), 33.

[‡] P. C. Clemmow, Quart. J. Mech. Appl. Maths., 3 (1950), 241.

[§] For a more detailed discussion of the method of stationary phase for single and double integrals see L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics* (Cambridge, Cambridge University Press, 1995), Sec. 3.3.

than in the form (1), and in practice the exponential commonly represents a travelling wave.

Keeping to the notation of (9), the paths of integration used are v(x, y) = constant; but in (14) this means that the amplitude part of the exponential is constant along the path, whilst the phase part varies most rapidly, a reversal of the situation in the method of steepest descent. It can still be established that the only significant contribution to the integral arises from portions of the path in the vicinity of saddle-points or end points, but the physical interpretation of the mechanism by which this comes about is now in terms of 'phase interference' (see §8.3, p. 429) rather than amplitude decay.

The method of stationary phase was first introduced explicitly by Lord Kelvin.* A rigorous mathematical treatment which would justify the statements made above was subsequently given by Watson;† this treatment is based on the fact that, if 0 < m < 1, a is a positive constant and F(x) has limited total fluctuation in the range $x \ge 0$, then

$$k^m \int_0^a x^{m-1} F(x) e^{ikx} dx \to F(0) \Gamma(m) e^{\frac{1}{2}im\pi}$$

as $k \to \infty$.‡

Watson's discussion, however, is rather restricted in scope. In particular, it does not seem capable of producing the complete asymptotic expansion. This is given in some detail by Focke§ for the case in which f(z) is a real function and the path of integration is confined to the real axis. Focke makes use of a neutralizing function, a method suggested earlier by van der Corput.¶

The deductions from the method of stationary phase follow much the same pattern as those from the method of steepest descent. For example, in the case when the path of integration in (14) starts from a saddle-point at z_0 and runs to infinity along v(x, y) = constant without encountering another saddle-point, the approximation corresponding to (12) is

$$\sqrt{-\frac{\pi}{2f''(z_0)}}g(z_0)e^{-\frac{1}{4}i\pi}\frac{e^{ikf(z_0)}}{\sqrt{k}}.$$
 (15)

But one aspect in which there is some distinction between the methods should be noted. With a steepest descent path which starts at a saddle-point and does not go to infinity, the contribution of the point at the end of the path to the strict asymptotic expansion is zero in comparison with the saddle-point contribution by virtue of the extra exponential factor it contains. On the other hand, with a stationary phase path of the same type the contribution of the point at the end of the path is, in general, of the order of that of the saddle-point merely divided by $k^{1/2}$; it is excluded, therefore, from the asymptotic approximation only if the first term of the asymptotic expansion is alone retained.

To sum up, the methods of steepest descent and stationary phase, stripped of their mathematical expression, depend on choosing a path of integration in such a way that

^{*} W. Thomson, Proc. Roy. Soc., A, 42 (1887), 80.

[†] G. N. Watson, Proc. Camb. Phil. Soc., 19 (1918), 49.

[‡] A result given by T. J. I'A. Bromwich, An Introduction to the Theory of Infinite Series (London, Macmillan, 1908), p. 447.

[§] J. Focke, Ber. Sächs. Ges. (Akad.) Wiss., 101 (1954), Heft 3.

[¶] J. G. van der Corput, *Indag. Math.*, **10** (1948), 201.

the integrand, by virtue of its exponential factor, contributes negligibly to the integral except in the vicinity of certain *critical points*, these being either saddle-points or end points of the path of integration.

3 Double integrals

It was shown in §8.3 and §9.1 that the problem of a field diffracted by an aperture leads to double integrals of the form

$$\iint g(x, y)e^{ikf(x,y)} dx dy, \tag{16}$$

where g(x, y) and f(x, y) are independent of k, and the domain of integration is determined by the aperture. Clearly (16) is analogous to (14), and approximations for large values of k are likewise provided by asymptotic expansions.

The theory of asymptotic expansions of double integrals is naturally more complicated than that of single integrals. The techniques of integration in the complex plane are only readily applicable to single integrals, so for double integrals it seems as though the attack must be based on an approach somewhat different from that developed in the preceding pages of this appendix.

The case in which f(x, y) is a real function is comprehensively discussed by Focke,* who uses the concept of a neutralizing function, in the paper already referred to in connection with the application of this idea to single integrals. The analysis again shows that contributions to the asymptotic expansion come only from regions in the vicinity of certain *critical points*, and that different types of critical point give rise to different powers of k in the leading terms of their respective contributions.†

There are three types of critical point. The leading terms of their respective contributions to the asymptotic expansion are now briefly discussed, excluding cases in which a critical point is of more than one type.

A critical point of the first kind is a point within the domain of integration at which

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial y} = 0. \tag{17}$$

Then near the critical point, (x_0, y_0) say,

$$f(x, y) = f(x_0, y_0) + \frac{1}{2}\alpha(x - x_0)^2 + \frac{1}{2}\beta(y - y_0)^2 + \gamma(x - x_0)(y - y_0) + \cdots, \quad (18)$$

where $\alpha = \partial^2 f/\partial x^2$, $\beta = \partial^2 f/\partial y^2$, $\gamma = \partial^2 f/\partial x \partial y$, the partial derivatives all being evaluated at (x_0, y_0) . Now choose certain new variables of integration ξ , η which are such that

$$f(x, y) = f(x_0, y_0) + \frac{1}{2}\alpha\xi^2 + \frac{1}{2}\beta\eta^2 + \gamma\xi\eta.$$
 (19)

^{*} J. Focke, loc. cit. See also D. S. Jones and M. Kline, J. Math. Phys., 37 (1958), 1; N. Chako, J. Inst. Maths Applics, 1 (1965), 372; M. Kline and I. W. Kay, Electromagnetic Theory and Geometrical Optics (New York, Interscience Publishers, 1965), Chapt. XII.

[†] Critical points for double integrals were briefly mentioned by J. G. van der Corput, *Indag. Math.*, **10** (1948), 201; and their qualitative significance for optics was pointed out by N. G. van Kampen, *Physica*, **14** (1949), 575.

Then the required asymptotic approximation to (16) is

$$g(x_0, y_0)e^{ikf(x_0, y_0)} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{\frac{1}{2}ik(\alpha\xi^2 + \beta\eta^2 + 2\gamma\xi\eta)} d\xi d\eta = \frac{2\pi i\sigma}{\sqrt{|\alpha\beta - \gamma^2|}} g(x_0, y_0) \frac{e^{ikf(x_0, y_0)}}{k},$$
(20)

where the positive square root* is taken and

$$\sigma = \begin{cases} +1 \text{ for } \alpha\beta > \gamma^2, \ \alpha > 0, \\ -1 \text{ for } \alpha\beta > \gamma^2, \ \alpha < 0, \\ -i \text{ for } \alpha\beta < \gamma^2. \end{cases}$$
 (21)

Expression (20) is the analogue for double integrals of the asymptotic approximation (15) for single integrals.

Critical points of the second kind are points on the curve bounding the domain of integration at which $\partial f/\partial s = 0$, where ds is an element of arc of the bounding curve. The power of k in the nonexponential part of the leading term of the corresponding contribution to the asymptotic expansion is $k^{-3/2}$, in contrast to the factor k^{-1} of (20).

Finally, *critical points of the third kind* are corner points on the curve bounding the domain of integration, that is, points at which the slope of the curve is discontinuous. In this case the corresponding factor is k^{-2} .

* The term under the square root sign has a simple geometrical interpretation. Consider the surface z = f(x, y). Let R_1 and R_2 be the principal radii of curvature and $K = 1/R_1R_2$ the Gaussian curvature at a typical point of the surface. Then [see, for example, G. Salmon, *Analytic Geometry of Three Dimensions*, Vol. I, revised by R. A. P. Rogers (London, Longmans, Green & Co., 5th edition, 1912), p. 411],

$$K = \frac{f_{xx}f_{yy} - f_{xy}^{2}}{(1 + f_{x}^{2} + f_{y}^{2})^{2}},$$

where $f_x = \partial f/\partial x$, $f_{xx} = \partial^2 f/\partial x^2$, etc. At a critical point of the first kind $f_x = f_y = 0$, $f_{xx} = \alpha$, etc., and this expression reduces to

$$K = \alpha\beta - \gamma^2$$
.