Stokes Phenomenon Demystified

R. B. PARIS, C. Math., FIMA
Department of Mathematical and Computer Sciences, University of Abertay Dundee
and A. D. WOOD, C. Math., FIMA
School of Mathematical Sciences, Dublin City University

1. Historical background

"THE inferior term enters as it were into a mist, is hidden for a little from view, and comes out with its coefficient changed", wrote the distinguished mathematical physicist Sir George Gabriel Stokes in a 1902 survey paper[1] at the end of his long and productive career. In such beautiful and poetic, but highly imprecise, language he sought to explain the phenomenon, observed by him for the Airy equation[2] in 1857 and for the Bessel equation[3] in 1868, of a change in the multiplier of one of the two linearly independent asymptotic solutions as a certain ray, which we now know as the Stokes line, in the complex plane is crossed. Perhaps as a boy growing up in his father's rectory at Skreen in County Sligo in the West of Ireland he had watched the mists skim the surface of flat-topped Benbulben across the bay, an area that was later to produce the poet, W. B. Yeats. Did this lead to an outbreak of poetry from this very private and reserved Victorian scientist? Whatever the reason, he left us with a deep and fascinating problem from which the mists have only properly cleared in the last five years, thanks to intense interest by physicists and mathematical analysts in the computation of exponentially small terms in asymptotic expansions.

This article seeks to resolve the paradox of why such apparently negligible terms can be of physical and computational importance in a variety of applications and reviews the theories, of varying degrees of rigour, which have been advanced in its support. The task of operating with divergent series, of which asymptotic series are a subset, has always been a delicate one. In his introduction to Hardy's book[4] on divergent series, Littlewood speaks of "the aroma of paradox and audacity which hangs about the subject." The credit for reviving interest in the Stokes phenomenon must go the the physicist M. V. Berry[5] who, in 1988, used Borel summation to resum the tail of the divergent asymptotic series, leading to the theories of *superasymptotics* and *hyperasymptotics* of Berry and Howls.[6] This in turn generated more rigorous work by mathematical analysts, including the *exponentially-improved asymptotic expansion* of Olver[7] and others. At the same time, a theory of *asymptotics beyond all orders* was being developed by the American applied mathematicians Kruskal and Segur.[8]

The definition of an asymptotic expansion had been put on a rigorous footing by Poincaré[9] in 1886, who defined the asymptotic power series expansion of a function $f(z)$ valid in some sector S of the complex plane by

$$f(z) \sim \sum_{r=0}^{\infty} a_r z^{-r} \quad \text{as} \quad |z| \to \infty \text{ in } S$$

if

$$z^n \left\{ f(z) - \sum_{r=0}^{n} a_r z^{-r} \right\} \to 0 \quad \text{as} \quad |z| \to \infty \text{ in } S.$$

In other words, the remainder in truncating the series at the $(n+1)$th term is of order of the first term omitted. This definition was extended to more general expressions involving an asymptotic scale $\phi_r(z)$ by setting

$$f(z) \sim g(z) \sum_{r=0}^{\infty} a_r \phi_r(z) \quad \text{as} \quad |z| \to \infty \text{ in } S$$

if

$$\left\{ f(z)/g(z) - \sum_{r=0}^{n} a_r \phi_r(z) \right\} \Big/ \phi_n(z) \to 0 \quad \text{as} \quad |z| \to \infty \text{ in } S.$$

The singular value need not always be at infinity, and expansions of the type

$$\phi(\epsilon) \sim \sum_{n=0}^{\infty} a_n \epsilon^n \quad \text{as} \quad \epsilon \to 0$$

frequently arise in perturbation theory. While the Poincaré definition has proved to be of enormous value in constituting a rigorous calculus for the manipulation of asymptotic expansions, it is not sufficiently refined to deal with problems which require a transcendental degree of precision. In the above perturbation-type expansion every term is algebraic in ϵ and transcendentally small terms like $\exp(-1/\epsilon^2)$, for example, are not captured under the Poincaré definition. Such terms are therefore said to be *asymptotic beyond all orders* in ϵ.

In the past it has been argued that in applications it must surely be safe to neglect such tiny correction terms, but it has been found recently, in a number of areas of physics, that this need not always be so. This arose initially in quantum mechanics in the paper of Pokrovskii and Khalatnikov on scattering from a potential barrier, in the independent work on adiabatic invariance of Meyer and Wasow, in extensive work on resonances by Simon, Harrell, Herbst and others, and recently in quantum chemistry with Connor and Brandas. The paper by Bender and Wu on the anharmonic oscillator should also be included in this group. More recently there has been extensive interest in the growth of dendritic crystals (Segur & Kruskal, McLeod & Amick, Levine) and in viscous fingering (Tanveer). Other physical applications include KAM theory for a rapidly forced pendulum (Scheurle, Marsden, Holmes) and solitary water waves (Beale, Shen & Sun, Byatt-Smith). The present authors were led to this field by working on a problem of energy leakage from a weakly-guiding optical fibre, first stated by Kath and Kriegsmann. The development of a mathematical theory to handle these

The Irish Branch of the IMA is arranging for the erection of a commemorative plaque at the birthplace of Stokes in Skreen. The event will take place on Friday 9th and Saturday 10th June, 1995 and will be accompanied by lectures on the life of Stokes and his mathematical contributions. Persons interested in receiving details should contact Professor Alastair Wood, School of Mathematical Studies, Dublin City University, Dublin 9. E-mail: wood@dcu.ie.

small exponentials clearly springs from the needs of physics and indeed, of industrial application in the fabrication of semiconductors and optical fibres. All references in this paragraph are to papers appearing or listed in the Proceedings of the Workshop on "Asymptotics Beyond All Orders", edited by Segur, Tanveer and Levine.[10]

We now provide, by way of example, three cases where a transcendentally small term, hiding behind algebraic terms of all orders in a divergent asymptotic series, can have a practical effect. Suppose the function $\phi(x, \epsilon)$ satisfies a differential equation in x which contains a small parameter ϵ. Suppose further that we have been able to construct a perturbation expansion

$$\phi(x, \epsilon) \sim \sum_{n=0}^{\infty} \phi_n(x)\epsilon^n \quad \text{as} \quad \epsilon \to 0, \text{ for all } x.$$

If each coefficient function $\phi_n(x)$ is symmetric in x about the origin, can we assert that the solution itself shares this property? Clearly not, since the presence of just one asymmetric transcendentally small term, which is excluded by the use of the Poincaré definition, is sufficient to destroy the symmetry. This mechanism is present in several of the physical examples above.

The second case arises in the theory of resonances and in optical tunnelling, where the quantity of physical relevance is the imaginary part of a complex eigenvalue of a differential equation boundary value problem, again involving a small parameter ϵ. It is known that the imaginary part of the eigenvalue is of a smaller order of magnitude than the real part. Indeed, in quantum mechanics, such resonance poles are an exponentially small distance off the negative real axis. In problems of this type we can obtain a perturbation expansion for the eigenvalue $\lambda(\epsilon)$ in powers of ϵ in which all the coefficients turn out to be real. How then do we proceed to compute the transcendentally small imaginary part of the eigenvalue?

Our third case concerns the use of the asymptotic series expansion of a function to compute values of the function for small, but moderate, values of the variable. This example is a slight modification of one on p. 76 of Olver.[11] Consider the function defined by the integral

$$I(\epsilon) = \int_0^{\pi} \frac{\cos(t/\epsilon)}{1+t^2}\,dt.$$

By means of integration by parts, it is easy to construct a series in powers of ϵ^2 which is asymptotic as ϵ tends to zero from above. Suppose we use this series to approximate the function at $\epsilon = 0{\cdot}1$. Retention of the first three terms in the expansion yields $I(0{\cdot}1) = -0{\cdot}0005271$, whereas direct numerical integration gives a value of $-0{\cdot}0004558$. This large relative error has arisen from neglect of an exponentially small term. This may be seen by rewriting the integral as the difference

$$I(\epsilon) = \tfrac{1}{2}\pi e^{-1/\epsilon} - \int_{\pi}^{\infty} \frac{\cos(t/\epsilon)}{1+t^2}\,dt,$$

where the first term results from routine evaluation round a semi-circular contour of the integral taken over $[0, \infty)$ and has the value $0{\cdot}0000713$ when $\epsilon = 0{\cdot}1$. Integration by parts of the second integral yields the same asymptotic power series as before, with the same value of $0{\cdot}0005271$. Together with the contribution from the exponentially small term, this provides the correct answer.

Having convinced the reader that there is a need for such precision asymptotics in a wide range of physical problems, the rest of this paper is devoted to dispersing the "mist" surrounding Stokes lines, thus enabling us to identify the correct exponentially small contribution to include in our computation in the neighbourhood of such lines. We shall take as our model the well-known modified Bessel function of the second kind, $K_\nu(z)$, as z tends to infinity. We first treat this in the spirit of Stokes' 1902 paper, halting at the problem of which expansion to choose in an overlapping region of validity about the Stokes line $\arg z = \pi$. We then outline the method used by Berry[5] to resolve this difficulty, before sketching the various rigorous treatments developed over the past five-year period.

2. The Stokes phenomenon

The Stokes phenomenon concerns the abrupt change across certain rays in the complex plane, known as Stokes lines, exhibited by the coefficients multiplying exponentially subdominant terms in compound asymptotic expansions. The root cause of this phenomenon appears to be a consequence of asymptotically approximating a given function, possessing a certain multivalued structure, in terms of approximants of a different multivalued structure. The simplest illustration of this for solutions of linear second order ordinary differential equations is Airy's equation

$$d^2y/dz^2 = zy, \tag{1}$$

for which the solutions $y(z)$ are entire functions of z. For large $|z|$, the asymptotic solutions to leading order are given by

$$u_\pm(z) = z^{-1/4}\exp\left(\pm\tfrac{2}{3}z^{3/2}\right),$$

which are multivalued functions with a branch point at $z = 0$. Consequently, any linear combination of these approximants $Cu_+(z) + Du_-(z)$, where C and D are arbitrary constants, cannot approximate a solution of (1) *uniformly* as one describes a complete circuit about $z = 0$, since $y(ze^{2\pi i}) = y(z)$ but $u_\pm(ze^{2\pi i}) \neq u_\pm(z)$. There must therefore be a change in the values of the constants C and D as $\arg z$ is increased by 2π.

A similar argument applies to the modified Bessel equation

$$z^2d^2y/dz^2 + zdy/dz - (z^2 + \nu^2)y = 0. \tag{2}$$

For non-integer ν, the linearly independent solutions of this equation in ascending powers of z are given by the modified Bessel functions of the first kind $I_{\pm\nu}(z)$, where

$$I_{\pm\nu}(z) = (z/2)^{\pm\nu}\sum_{k=0}^{\infty}\frac{(z/2)^{2k}}{k!\,\Gamma(k\pm\nu+1)} \quad (|z| < \infty),$$

which are associated with the indicial exponents $\pm\nu$ at the branch point $z = 0$. The (formal) asymptotic solutions for large z and fixed real or complex values of the parameter ν are given by

$$\begin{aligned} u_+(z) &= \left(\frac{\pi}{2z}\right)^{1/2} e^{z}\sum_{r=0}^{\infty}(-)^r A_r(\nu)(2z)^{-r},\\ u_-(z) &= \left(\frac{\pi}{2z}\right)^{1/2} e^{-z}\sum_{r=0}^{\infty} A_r(\nu)(2z)^{-r}, \end{aligned} \tag{3}$$

where $A_0(\nu) = 1$ and

$$A_r(\nu) = \frac{(4\nu^2 - 1^2)(4\nu^2 - 3^2)\dots(4\nu^2 - (2r-1)^2)}{r!\,2^{2r}} \qquad (r \geq 1) \quad (4)$$

$$= (-)^r \frac{\cos \pi\nu}{\pi} \frac{\Gamma(r + \nu + \frac{1}{2})\Gamma(r - \nu + \frac{1}{2})}{r!}. \qquad (5)$$

Again (provided 2ν is not an odd integer), a Stokes phenomenon has to occur whenever a solution of (2) in the form $y(z) = A'I_\nu(z) + B'I_{-\nu}(z)$, where A', B' are arbitrary constants, is expressed in terms of the linear combination $Cu_+(z) + Du_-(z)$ for large $|z|$, since $y(z)$ and its asymptotic approximants $u_\pm(z)$ have a different multivalued structure. This qualitative argument that the asymptotic approximation must be *domain-dependent* was first given by Stokes;[1] for a fuller account, see the review paper by Meyer.[12] Stokes went on to reason that the change in the constants C, D must take place where the terms of the associated expansions $u_\pm(z)$ come to be regularly positive; this is equivalent to the rays where $u_\pm(z)$ assumes maximal dominance over $u_\mp(z)$ for large $|z|$. Thus the values of $\theta = \arg z$ across which C and D *can* change are given by $\theta = (2n+1)\pi$ and $\theta = 2n\pi$ ($n = 0, \pm 1, \pm 2, \dots$), respectively.

To be more specific we now focus our attention on the modified Bessel function $K_\nu(z)$ which is defined to be that combination of $I_{\pm\nu}(z)$ (corresponding to $-A' = B' = \frac{1}{2}\pi \operatorname{cosec} \pi\nu$) which yields $C = 0$, $D = 1$ in the sector $-\pi < \theta < \pi$, that is

$$K_\nu(z) \sim u_-(z), \quad |z| \to \infty, \quad -\pi < \theta < \pi. \qquad (6)$$

The Stokes lines for $K_\nu(z)$ are consequently situated at $\theta = \pm\pi, \pm 2\pi, \dots$. In the following discussion we shall consider only the Stokes line $\theta = \pi$ across which C must change; the treatment of the other rays is similar. In the adjacent sector $\pi < \theta < 2\pi$ the value of C can be computed directly from the branch point structure of the convergent solutions $I_{\pm\nu}(z)$, or equivalently by use of the connection formula

$$K_\nu(z) = 2\cos \pi\nu\, K_\nu(ze^{-\pi i}) - K_\nu(ze^{-2\pi i}).$$

On replacing z in (6) by $ze^{-\pi i}$ and $ze^{-2\pi i}$ in turn, we obtain the compound asymptotic expansion

$$K_\nu(z) \sim u_-(z) + 2i \cos \pi\nu\, u_+(z) \qquad |z| \to \infty, \quad \pi < \theta < 2\pi. \quad (7)$$

Thus, on crossing $\theta = \pi$ in the positive sense with $|z|$ fixed, the constant C multiplying the subdominant asymptotic solution $u_+(z)$ changes from zero to the value $2i \cos \pi\nu$.

It is worth pointing out that Stokes initially obtained the change in the subdominant coefficient by a more involved argument which made use of the asymptotic behaviour along certain rays of a suitable integral representation of the function in question. In addition, he supported his theory by a numerical example (applied to the Airy function $Ai(z)$) in which he compared the predictions from the different asymptotic expansions either side of a Stokes line with the exact values. To detect the "birth" of such exponentially small terms in this way necessitates the computation of the dominant series to at least a comparable precision. Stokes achieved this by employing the by-now familiar process of optimal truncation of the dominant asymptotic expansion (that is, truncation just before the least term of the expansion) together with an estimate of the associated remainder term, thereby anticipating Stieltjes[13] by about 30 years.

In the literature the above asymptotic expansions for $K_\nu(z)$ are normally stated in the less stringent sense of Poincaré, namely

$$K_\nu(z) \sim u_-(z) \qquad |\theta| \leq \tfrac{3}{2}\pi - \delta \qquad (8)$$

$$K_\nu(z) \sim u_-(z) + 2i \cos \pi\nu\, u_+(z) \qquad \tfrac{1}{2}\pi + \delta \leq \theta \leq \tfrac{5}{2}\pi - \delta, \qquad (9)$$

where throughout δ denotes an arbitrarily small positive constant. These expansions hold uniformly as $|z| \to \infty$ with respect to $\theta = \arg z$ in their respective sectors of validity. When 2ν is not an odd integer, the expansions (8) and (9) have the common sector of validity, $\frac{1}{2}\pi + \delta \leq \theta \leq \frac{3}{2}\pi - \delta$, where they differ by the inclusion of the series $u_+(z)$. The resulting contradiction, however, is only apparent, since $u_+(z)$ is uniformly exponentially small for large $|z|$ in this common sector and so is asymptotically smaller than any term in the dominant series. In the Poincaré sense, therefore, the additional term in (9) is negligible.

Since Stokes's discovery of this phenomenon over 130 years ago, the conventional view has been that of a discontinuous change in the constants (called *multipliers*) associated with subdominant asymptotic expansions. The fact that an analytic function can possess such a discontinuous representation in the neighbourhood of an irregular singular point has always been problematic. This "hide-and-seek" nature of the multipliers, together with the inherent vagueness associated with the precise location of these jumps in the complex plane, has resulted in the Stokes phenomenon being enshrouded in a certain element of mystery. In 1980 Olver,[14] in a survey paper dealing with well-constructed error bounds for asymptotic approximations, re-examined the question of which expansion in (8) and (9) to adopt for $K_\nu(z)$ in the common sector of validity, $\frac{1}{2}\pi + \delta \leq \theta \leq \frac{3}{2}\pi - \delta$. By a detailed analysis of the associated error bounds, he concluded that, apart from a region of uncertainty R bounded by a truncated parabola with axis along the negative real z-axis, the expansion (8) is more accurate when $\theta < \pi$ while the expansion (9) is more accurate when $\theta > \pi$. Inside the region R, the error bounds were insufficiently precise to resolve the details of the transition from one expansion to the other. However, Olver was able to deduce that $\theta = \pi$ should be taken as the Stokes line for $K_\nu(z)$, since this is the only ray that is completely contained in R.

3. Smoothing of the Stokes phenomenon

This picture changed dramatically when Berry[5] argued that the coefficient of the subdominant expansion should be regarded not as a discontinuous constant but, for fixed $|z|$, as a continuous function of θ. Berry's theory was based on Dingle's earlier interpretative theory[15] of asymptotic expansions and demonstrated that, when viewed on an appropriately magnified scale, the change in the subdominant multiplier near a Stokes line is in fact *continuous*. For a wide class of functions, the functional form of this rapid but smooth transition is found to possess a universal structure represented to leading order by an error function, whose argument is an appropriate variable (dependent on the function in question) describing the transition across a Stokes line.

To summarise this theory we continue with our example of $K_\nu(z)$ and optimally truncate $u_-(z)$ in (6) just before its least term. This is given roughly by $2|z| \simeq |A_N(\nu)/A_{N-1}(\nu)|$ which, from (4), corresponds to

$$N - 1 = [2|z|], \tag{10}$$

where [ ] denotes the integer part. The Stokes multiplier S_N is then *defined* by the equation

$$K_\nu(z) = \left(\frac{\pi}{2z}\right)^{1/2} e^{-z} \sum_{r=0}^{N-1} A_r(\nu)(2z)^{-r} + 2i\cos\pi\nu\left(\frac{\pi}{2z}\right)^{1/2} e^{z} S_N, \tag{11}$$

which, for fixed $|z|$ (and hence fixed N), is a continuous function of θ. Since the term containing S_N is proportional to the leading term of the expansion of $u_+(z)$, we see from (6) and (7) that as we traverse the Stokes line $\theta = \pi$ (with fixed $|z| \to \infty$) the value of S_N must change from 0 to 1.

Berry's argument was that the second term on the right-hand side of (11) originates from a correct interpretation of the tail of the divergent expansion (6). By comparison of (6) with (11), he formally identified the Stokes multiplier in terms of the remainder of the expansion $u_-(z)$ truncated at the Nth term as

$$S_N \equiv \frac{e^{-2z}}{2i\cos\pi\nu} \sum_{r=N}^{\infty} A_r(\nu)(2z)^{-r}.$$

For large N, the coefficients $A_r(\nu)$ $(r \geqslant N)$ can be approximated by use of Stirling's formula to reduce the quotient of gamma functions in (5) to a single gamma function, namely

$$A_r(\nu) \sim (-)^r \frac{\cos\pi\nu}{\pi}\Gamma(r), \quad r \geqslant N \gg 1,$$

so that

$$S_N \simeq \frac{e^{-2z}}{2\pi i} \sum_{r=N}^{\infty} \Gamma(r)(2ze^{-\pi i})^{-r}. \tag{12}$$

The series in (12) is, of course, divergent and so numerically meaningless but Berry showed that, if correctly interpreted by Borel summation, it can lead to a *convergent* representation for S_N from which the structure of the multiplier in the vicinity of the Stokes line can be deduced. This is achieved by replacing $\Gamma(r)$ in (12) by its well-known Euler integral representation and interchanging the order of summation and integration to yield

$$S_N \simeq \frac{e^{-2z}}{2\pi i} \sum_{r=N}^{\infty} (2ze^{-\pi i})^{-r} \int_0^\infty e^{-t} t^{r-1}\, dt = \frac{e^{-2z}}{2\pi i} \int_C \tau^{-1} e^{2z\tau} \sum_{r=N}^{\infty} \tau^r\, d\tau,$$

where we have put $\tau = e^{\pi i} t/2z$ and C denotes the ray in the τ-plane specified by $[0, \infty \exp(\pi - \theta)i)$ (see Fig. 1). Straightforward formal evaluation of the series as $\tau^N/(1-\tau)$ then finally yields the convergent ap-

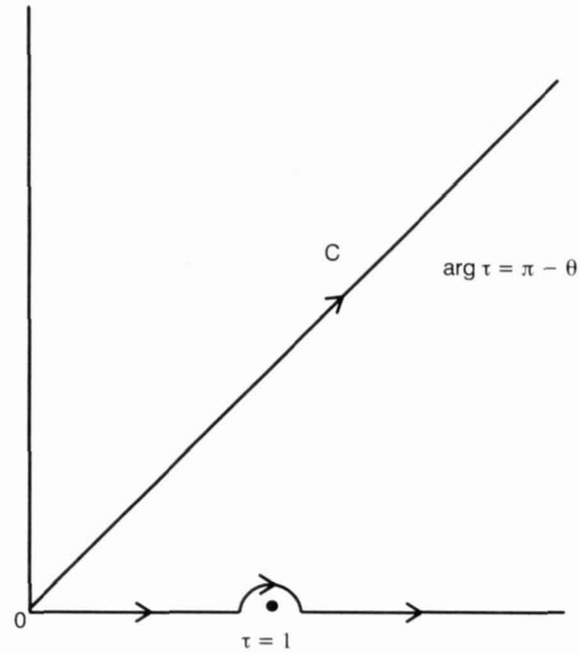

Fig. 1. The path of integration in the τ plane and the indentation round the pole at $\tau = 1$

proximation for the Stokes multiplier in the form

$$S_N \simeq \frac{1}{2\pi i} \int_C e^{-2z(1-\tau)} \frac{\tau^{N-1}}{1-\tau}\, d\tau. \tag{13}$$

This integral can be converted into a more manageable form by means of Cauchy's theorem, provided $\frac{1}{2}\pi < \theta < \frac{3}{2}\pi$, to swing round the path of integration on to the positive real axis, together with a semicircular indentation round the pole at $\tau = 1$, as shown in Fig. 1. Thus

$$S_N \simeq \tfrac{1}{2} + \frac{1}{2\pi i} P\int_0^\infty e^{-2z(1-\tau)} \frac{\tau^{N-1}}{1-\tau}\, d\tau, \quad (\tfrac{1}{2}\pi < \theta < \tfrac{3}{2}\pi)$$

where the $\frac{1}{2}$ results from the contribution round the indentation as its radius tends to zero and P denotes the Cauchy principal value. The characteristic feature of this integral, which dominates its behaviour for large $|z|$, is the proximity of the saddle point at $\tau \simeq e^{(\pi-\theta)i}$ (when N is optimal) to the pole at $\tau = 1$ when $\theta \simeq \pi$. We therefore put

$$2ze^{-\pi i} = A + iB, \quad N - 1 = A + \mu, \tag{14}$$

where A, B are real, and restrict our attention to a neighbourhood of $\theta = \pi$ so that $A \gg 1$ and $|B| \ll A$. The region near the Stokes line is made more specific if, in addition, we stipulate that μ is of order unity; for this requires that $|B|/\sqrt{2A}$ be bounded since, from (10), $N - 1 = [\sqrt{A^2 + B^2}] = [A + B^2/2A + \ldots]$. (We remark that this region is similar to Olver's region of uncertainty R.)

Defining the new variable $\tau = 1 + u$ and $f(u) = u - \ln(1+u)$, we then obtain

$$S_N \simeq \tfrac{1}{2} - \frac{1}{2\pi i} P\int_{-1}^\infty u^{-1} e^{-Af(u)} e^{-iBu}(1+u)^\mu\, du.$$

Since the dominant behaviour for $A \to +\infty$ will arise from the neighbourhood of $u = 0$, we replace $f(u)$ and (since μ is bounded) $(1+u)^\mu$ by the first term of their respective Maclaurin expansions and extend the lower limit of integration to $-\infty$. In this way we find approximately

$$S_N \simeq \tfrac{1}{2} - \frac{1}{2\pi i} P \int_{-\infty}^{\infty} u^{-1} e^{-Au^2/2} e^{-iBu}\, du$$
$$= \tfrac{1}{2} - \frac{1}{\pi} \int_0^{\infty} e^{-Au^2/2} \frac{\sin Bu}{u}\, du,$$

where only the even part contributes to the principal value of the integral.

Evaluation of this last integral[16] then finally yields for $|z| \to \infty$

$$S_N \simeq \tfrac{1}{2} + \tfrac{1}{2}\,\mathrm{erf}\,\sigma, \quad \sigma = \frac{B}{\sqrt{2A}}, \tag{15}$$

where erf denotes the error function defined by

$$\mathrm{erf}\,x = \frac{2}{\sqrt{\pi}} \int_0^x e^{-u^2}\, du.$$

This result indicates the functional form of the Stokes multiplier for $K_\nu(z)$ in the vicinity of the Stokes line $\theta = \pi$. The change is expressed in terms of the variable σ which provides a *local* measure of the angular transition across the Stokes line at constant large $|z|$. We note that the error function for real argument lies between the values ± 1 and that these limiting values are approached rapidly; for example, $|\mathrm{erf}\,(\pm 2)| > 0{\cdot}995$. As a consequence, S_N increases smoothly but rapidly from the value 0 when σ is moderately large (but bounded) and negative (*i.e.*, when θ is somewhat less than π) to the value 1 when σ is moderately large and positive (*i.e.*, when θ is somewhat greater than π). On the Stokes line we find $S_N \simeq \tfrac{1}{2}$.

From (12) we see that the coefficients in the tail of the optimally truncated expansion $u_-(z)$ behave to leading order like a "factorial divided by a power of z". Berry developed his theory for a very wide class of functions for which the coefficients in the tails of their asymptotic expansions possess a similar structure. The representation of S_N as a convergent integral and its subsequent estimation are similar to that for $K_\nu(z)$ and are expressed in terms of the "singulant" F. This quantity was defined by Dingle[15] as the difference between the exponents of the dominant and subdominant exponentials (in the case of $K_\nu(z)$, we have $F = -2z$; cf. (14)). The behaviour of S_N near a Stokes line for this more general class of functions is similarly found to possess the error function dependence given by (15), but with the variable σ now given locally by

$$\sigma = \mathrm{Im}\,(F)/(2\,\mathrm{Re}\,(F))^{1/2}. \tag{16}$$

4. Rigorous approaches

The arguments put forward by Berry[5] were highly innovative and insightful but were nevertheless quite formal. In addition, the universal error function dependence only describes the leading behaviour of the Stokes multiplier locally in a roughly parabolic-shaped domain about the Stokes line (corresponding to bounded values of σ). Since the appearance of this seminal paper, several authors have placed this valuable theory on a rigorous foundation and have established more refined approximations to S_N.

Olver,[17,18] following on from his discussion of the estimation of remainder terms in the last chapter of his book,[11] was the first to construct new uniform exponentially-improved asymptotic expansions for functions defined by Laplace integrals. Such generalised expansions have the merit of possessing greater accuracy and a larger domain of validity than conventional Poincaré-type expansions. To illustrate his theory, he chose the confluent hypergeometric function $U(a; a - b + 1; z)$, which includes many of the commonly used special functions, such as Bessel, Airy and parabolic cylinder functions. This function is defined by

$$U(a; a-b+1; z) = \frac{1}{\Gamma(a)} \int_0^{\infty} e^{-zt} t^{a-1} (1+t)^{-b}\, dt, \qquad \mathrm{Re}\,(a) > 0, \tag{17}$$

in the sector $|\theta| < \tfrac{1}{2}\pi$, and, like $K_\nu(z)$, possesses Stokes lines at $\theta = \pm n\pi$, $n = 1, 2, \ldots$. This is not surprising since $K_\nu(z) = \sqrt{\pi}(2z)^\nu e^{-z} U(\nu + \tfrac{1}{2}; 2\nu + 1; 2z)$. The asymptotic expansions in the Poincaré sense as $|z| \to \infty$ are

$$U(a; a-b+1; z) \sim u_-(z) \quad |\theta| \leq \tfrac{3}{2}\pi - \delta$$
$$U(a; a-b+1; z) \sim u_-(z) - \frac{2\pi i e^{-(a+b)\pi i}}{\Gamma(a)\Gamma(b)} u_+(z),$$
$$\tfrac{1}{2}\pi + \delta \leq \theta \leq \tfrac{5}{2}\pi - \delta$$

where

$$u_+(z) = z^{b-1} e^z \sum_{r=0}^{\infty} \frac{(1-a)_r (1-b)_r}{r!\, z^r},$$
$$u_-(z) = z^{-a} \sum_{r=0}^{\infty} (-)^r \frac{(a)_r (b)_r}{r!\, z^r}$$

and $(a)_r = \Gamma(a+r)/\Gamma(a)$. For the Stokes line $\theta = \pi$, the Stokes multiplier S_N is then defined as in (11) by

$$U(a; a-b+1; z) = z^{-a} \sum_{r=0}^{N-1} (-)^r \frac{(a)_r (b)_r}{r!\, z^r} - \frac{2\pi i e^{-(a+b)\pi i}}{\Gamma(a)\Gamma(b)} z^{b-1} e^z S_N, \tag{18}$$

where the dominant series $u_-(z)$ is optimally truncated at the Nth term, given by $N = [|z| - \mathrm{Re}\,(\alpha) - 1]$, $\alpha = a + b - 1$. It then follows that, for fixed large $|z|$, S_N increases from 0 to 1 as θ increases from values just below π to values just above π.

To obtain a representation for S_N valid in a domain containing the Stokes line, Olver first analytically continued the integral (17) (by suitable rotation of the path of integration) followed by expansion of the factor $(1+t)^{-b}$ into a series truncated after N terms together with a remainder term. In this way he was able to express S_N exactly in the form of a double integral in a half-plane containing $\theta = \pi$. This integral was then shown to possess the expansion for $|z| \to \infty$, valid in the sector $\tfrac{1}{2}\pi - \gamma < \theta < \tfrac{3}{2}\pi - \gamma$, (with $0 < \gamma \leq \tfrac{1}{2}\pi$),

$$S_N \sim \frac{1}{2\pi i} \sum_{r=0}^{\infty} \frac{(1-a)_r (1-b)_r}{r!\, z^r} \times \int_{C'} e^{-z(1-\tau)} \frac{\tau^{N+\alpha-r-1}}{1-\tau}\, d\tau, \tag{19}$$

where C' denotes the ray $[0, \infty e^{i\gamma})$ similar to that shown in Fig. 1.

The integrals appearing in the above sum are analogous to that in (13) and can be expressed in terms of the incomplete gamma function $\Gamma(1-\nu, z)$, which appears as a "building block" in the theory of the Stokes phenomenon. This is achieved by introduction of Dingle's so-called "terminant" function $T_\nu(z)$ defined by

$$T_\nu(z) = \frac{e^{\pi\nu i}\Gamma(\nu)}{2\pi i}\Gamma(1-\nu, z)$$
$$= \frac{1}{2\pi i}\int_{C'} e^{-z(1-\tau)}\frac{\tau^{\nu-1}}{1-\tau}d\tau, \quad \mathrm{Re}\,(\nu) > 0$$

where the path C' and the sector of validity of the integral are the same as in (19). For $\nu \sim |z| \gg 1$, the asymptotic behaviour of this integral is again controlled by a saddle point at $\tau = -e^{-i\theta}$ which becomes coincident with the pole of the integrand at $\tau = 1$ when $\theta = \pm\pi$. Olver[7] has shown that $T_\nu(z)$ has the following asymptotic expansion for $\nu \sim |z| \to \infty$ uniformly valid in θ:

$$T_\nu(z) \sim \tfrac{1}{2} + \tfrac{1}{2}\,\mathrm{erf}\,\{c(\theta)\sqrt{\tfrac{1}{2}|z|}\} - \frac{ie^{-(1/2)|z|c^2(\theta)}}{\sqrt{2\pi|z|}}\sum_{r=0}^{\infty} a_r|z|^{-r},$$
$$-\pi + \delta \leqslant \theta \leqslant 3\pi - \delta$$

where the a_r are calculable coefficients which depend on θ (and on $|z| - \nu$). (A conjugate expansion is valid in the sector $-3\pi + \delta \leqslant \theta \leqslant \pi - \delta$.) The quantity $c(\theta)$, which measures the proximity of the saddle point to the pole, is defined as that branch of $\frac{1}{2}c^2(\theta) = 1 + i(\theta - \pi) - e^{i(\theta-\pi)}$ that behaves in the vicinity of $\theta = \pi$ like

$$c(\theta) = \theta - \pi + \tfrac{1}{6}i(\theta-\pi)^2 - \tfrac{1}{36}(\theta-\pi)^3 + \ldots.$$

In this fashion Olver obtained from (19) by rigorous analysis the result for $|z| \to \infty$

$$S_N \sim \sum_{r=0}^{\infty}\frac{(1-a)_r(1-b)_r}{r!\,z^r}T_{N+\alpha-r}(z) \qquad (20)$$

uniformly valid with respect to θ in the sector $\frac{1}{2}\pi + \delta \leqslant \theta \leqslant \frac{3}{2}\pi - \delta$. It is then possible to construct the uniform exponentially-improved asymptotic expansion of S_N from knowledge of the expansion of $T_\nu(z)$. The leading order approximation to S_N, given by $r = 0$, is therefore

$$S_N = T_{N+\alpha}(z) + O(z^{-1})$$
$$= \tfrac{1}{2} + \tfrac{1}{2}\,\mathrm{erf}\,\{c(\theta)\sqrt{\tfrac{1}{2}|z|}\} + e^{-(1/2)|z|c^2(\theta)}O(|z|^{-1/2}), \quad (21)$$

which is seen to be independent of the parameters a and b. Near $\theta = \pi$, the argument of the error function is approximately $(\theta - \pi)\sqrt{|z|/2}$, which to leading order is equivalent to Berry's local variable σ in (16) (when $F = -z$). This shows that for large $|z|$, S_N changes smoothly in a zone of angular width about $\theta = \pi$ scaling like $|z|^{-1/2}$. The result (21), however, reveals how the error function argument is modified beyond the domain where σ is bounded.

Independently of both Berry and Olver, Jones[19] considered the asymptotic behaviour of the optimally truncated remainder terms for integrals of type

$$\int_0^1 f(u)e^{-zu}\,du.$$

By expressing $f(u)$ as an inverse Laplace transform, he reduced this integral to a Stieltjes integral of the form

$$\int_0^\infty \frac{F(t)e^{-t}}{z+t}dt$$

and, although the question of the Stokes phenomenon was not addressed, he established that the optimal remainder has behaviour of error function type near $\theta = \pm\pi$. Subsequently, Boyd[20] developed an exponentially-improved asymptotic theory for functions defined by a Stieltjes transform, choosing as an example to illustrate his theory the modified Bessel function $K_\nu(z)$. Starting from the expression

$$K_\nu(z) = z^{-1/2}e^{-z}\frac{\cos\pi\nu}{\pi}$$
$$\times\int_0^\infty \frac{e^{-t}t^{-1/2}}{1+t/z}K_\nu(t)\,dt, \quad (z \neq 0)$$

defined for $|\mathrm{Re}\,(\nu)| < \frac{1}{2}$ and $|\theta| < \pi$, he expanded the denominator $(1 + t/z)^{-1}$ into a series truncated at its Nth term together with a remainder. Upon term-by-term integration, this yields the result (11) with the Stokes multiplier now specified by

$$S_N = -\frac{1}{2\pi}\left(\frac{\pi}{2z}\right)^{-1/2}e^{-2z}\int_C e^{z\tau}\frac{\tau^{N-1/2}}{1-\tau}K_\nu(z\tau e^{-\pi i})\,d\tau,$$

where $\tau = e^{\pi i}t/z$ and C denotes the same path of integration as in Fig. 1.

Since the integrand is dominated by its behaviour near $\tau = 1$ (where the argument of $K_\nu(z\tau e^{-\pi i})$ is large and almost positive when $\theta \simeq \pi$), Boyd replaced the Bessel function by its dominant expansion given in (6). In this way, by term-by-term integration combined with analytical continuation arguments to extend the domain of validity in θ (and in ν), he was able to establish the asymptotic expansion of the Stokes multiplier for $K_\nu(z)$ for large $|z|$ in the form

$$S_N \sim \frac{1}{2\pi i}\sum_{r=0}^{\infty}\frac{A_r(\nu)}{(-2z)^r}\int_C e^{-2z(1-\tau)}\frac{\tau^{N-r-1}}{1-\tau}d\tau \qquad (22)$$
$$= \sum_{r=0}^{\infty}\frac{A_r(\nu)}{(-2z)^r}T_{N-r}(2z) \qquad (23)$$

in the domain $\frac{1}{2}\pi \leqslant \theta \leqslant \frac{3}{2}\pi$ containing the Stokes line $\theta = \pi$. The leading term of (22) corresponds to Berry's result in (13). Also, we see that the expansion (23) agrees with Olver's result for the confluent hypergeometric function in (20) when the values $a = \frac{1}{2} + \nu$, $b = \frac{1}{2} - \nu$ are inserted and z is replaced by $2z$.

Shortly after this, Paris[21] adopted a different viewpoint and considered functions defined by Mellin-Barnes integrals. These integrals usually involve one or more gamma functions (with possibly trigonometric functions) and their associated flexibility, combined with the properties of the gamma function, readily enable the construction of uniform exponentially-improved asymptotic expansions. Again this theory was illustrated by reference to the confluent hypergeometric function, for which

$$U(a; a+b-1; z) = \frac{z^{-a}}{2\pi i\Gamma(a)\Gamma(b)}$$
$$\times\int_{L_0}\Gamma(-s)\Gamma(a+s)\Gamma(b+s)z^{-s}\,ds, \quad |\theta| < \tfrac{3}{2}\pi \quad (24)$$

where L_n $(n = 0, 1, 2, \ldots)$ denotes the path $(c + n - \infty i, c + n + \infty i)$ parallel to the imaginary s axis, with $0 < c < 1$. (The path L_0 may be indented, if necessary, to separate the poles of $\Gamma(-s)$ from those of $\Gamma(a+s)$ and $\Gamma(b+s)$.) This representation has the advantage of defining $U(a; a+b-1; z)$ in the wider sector $|\theta| < \frac{3}{2}\pi$, so that analytical continuation arguments, which are essential for both Laplace and Stieltjes transforms, are unnecessary for the Stokes lines at $\theta = \pm\pi$.

Straightforward displacement of the path of integration in (24) to the right and evaluation of the residues of the poles of $\Gamma(-s)$ at $s = 0, 1, \ldots, N-1$ then leads to the result in (18), with the Stokes multiplier given by

$$S_N = \frac{(e^{-\pi i}z)^{-\alpha}e^{-z}}{4\pi}\int_{L_N}\frac{\Gamma(a+s)\Gamma(b+s)}{\Gamma(1+s)}\frac{z^{-s}}{\sin \pi s}\,ds,$$

valid throughout the sector $|\theta| < \frac{3}{2}\pi$, and where we recall that $\alpha = a+b-1$. The approximation of this integral for large $|z|$ to yield the expansion given in (20) then immediately follows upon expansion of the quotient of gamma functions on the path L_N as a series of inverse factorials in the form

$$\frac{\Gamma(a+s)\Gamma(b+s)}{\Gamma(1+s)} \sim \sum_{r=0}^{\infty}(-)^r\frac{(1-a)_r(1-b)_r}{r!} \times\Gamma(\alpha+s-r) \quad |s|\to\infty, \quad |\arg s| < \pi$$

and identification of the resulting integrals in terms of the terminant function by means of the Mellin-Barnes representation

$$-2iT_{N+\beta}(z) = \frac{(e^{-\pi i}z)^{-\beta}e^{-z}}{2\pi i} \times\int_{L_N}\Gamma(\beta+s)\frac{z^{-s}}{\sin \pi s}\,ds, \quad |\theta| < \tfrac{3}{2}\pi.$$

All these theories are quite general and are applicable to a wide class of functions defined by an integral transform of the corresponding type. Whichever integral representation approach is taken, however, the conclusion is the same: the use of the asymptotic approximations (6), (7) by themselves is unsafe in the neighbourhood of the Stokes line at $\theta = \pi$. As $\theta \to \pi$ from below, the contribution of the series $u_+(z)$ must increasingly be taken into account in computation, even though (7) says that it does not appear until $\theta > \pi$. On the Stokes line itself, the correct approximation is the average of (6) and (7).

5. Recent developments

It is important to understand the fundamental difference in philosophy between working directly with the tail of the asymptotic series and adopting an integral representation approach. It is true, as Dingle points out in his book,[15] that the tail of the divergent series contains information which may be decoded to give the form of the dominant expansion in the adjacent sector. In the case when the terms are of the form of a factorial divided by a power of z, the gamma function representation is used to replace the factorial by an integral. The non-rigorous method of Borel summation is then applied to interchange summation and integration to obtain a single integral for the remainder term. This integral is manipulated to eventually become the terminant. In the integral representation approach, we start from the other end of the problem and such a difficulty does not arise, making it possible to give a rigorous proof on the lines of §4.

Early in 1991 it was felt that a third approach should be possible. Based on the fact that all functions considered up to that date were solutions of some second order linear differential equation, it should be possible to obtain the exponentially-improved asymptotic solution directly from the equation without prior knowledge of the divergent series or an integral representation of the solution. Again this was done non-rigorously by Berry[22] and shortly afterwards by McLeod[23] for a special class of equation. A proof for the confluent hypergeometric equation was given by Olver[24] in 1992, and this has recently been generalised to equations with irregular singularity of rank 1 by Olver and Olde Daalhuis.[25] A class of equations of arbitrary order n with power coefficients has been treated by Paris[26] in 1992 using the Mellin transform approach: with more than two linearly independent solutions, the question of subdominance becomes more difficult. In this case there can be up to n distinct exponential behaviours at infinity to consider.

Not all special functions of interest in applications satisfy a second order linear equation. Such a function is the gamma function, which was shown in the nineteenth century to be incapable of satisfying a linear ordinary differential equation with rational coefficients. Starting from the logarithm of the gamma function and using a Mellin-Barnes integral representation involving the Riemann zeta function, the present authors[27] obtained the exponentially-improved expansion for the gamma function in 1990. They observed that infinitely many exponential behaviours each associated with its own Stokes multiplier were possible at infinity, but the detailed computation of these was subsequently carried out by Berry.[28] Although the gamma function does not satisfy a differential equation, it does satisfy a difference equation. Given that a difference equation may be regarded as a differential equation of infinite order, this explains why there are infinitely many exponential behaviours at infinity for the gamma function. Recently, Lawless[29] obtained exponentially-improved asymptotics for solutions of first order difference equations with rational function coefficients: this included the gamma function as a special case.

Another extension of exponential asymptotics has been the theory of *hyperasymptotics*, the theory which we described in §3 being known as *superasymptotics*. This nomenclature is due to Berry and Howls, who carried out formal explorations,[6] although the work of Boyd[20] slightly predates this. More recent rigorous justification has been provided by Olde Daalhuis.[30] Recall from §4 that the remainder S_N in optimally truncating the dominant series $u_-(z)$ was given by the asymptotic series (20) or (23). Their basic idea was to again optimally truncate either (20) or (23) and estimate the new remainder asymptotically, thus picking up terms of exponentially smaller order. This process may be repeated until the desired level of precision is achieved. A hyperasymptotic extension of the saddle point method has also been described by Berry and Howls[31] and Howls.[32]

It is important to realise that all of this work has taken place over a five-year period, after the topic of the Stokes phenomenon had lain virtually unnoticed, except by practitioners in the field, for the better part of a century. Intense concentration on this "hot topic in applied mathematics" has dispelled Stokes' mist to reveal that the multiplier does not jump across a Stokes line, but changes rapidly and continuously in a generic manner, turning on its head standard teaching in mathematical methods courses in this century. But, more than that, it is hoped that the rigorous work by the mathematical analysts has removed the last whiff of paradox and audacity from this fascinating subject.

References

1. Stokes, G. G., "On the discontinuity of arbitrary constants that appear as multipliers of semi-convergent series." *Acta Mathematica*, 1902, **26,** 393–397. Reprinted in "Mathematical and Physical Papers," Cambridge University Press, 1905, vol. 5, pp. 283–287.
2. Stokes, G. G., "On the discontinuity of arbitrary constants which appear in divergent developments." *Trans. Camb. Phil. Soc.*, 1857, **10,** 106–128. Reprinted in "Mathematical and Physical Papers," Cambridge University Press, 1905, vol. 4, pp. 77–109.
3. Stokes, G. G., "Supplement to a paper on the discontinuity of arbitrary constants which appear in divergent developments." *Trans. Camb. Phil. Soc.*, 1868, **11,** 412–423. Reprinted in "Mathematical and Physical Papers," Cambridge University Press, 1905, vol. 4, pp. 283–298.
4. Hardy, G. H., "Divergent Series," Oxford University Press, 1949.
5. Berry, M. V., *Proc. Roy. Soc. Lond.*, 1989, **A422,** 7–21.
6. Berry, M. V. and Howls, C. J., *Proc. Roy. Soc. Lond.*, 1990, **A430,** 653–668.
7. Olver, F. W. J., *SIAM J. Math. Anal.*, 1991, **22,** 1460–1489.
8. Kruskal, M. D. and Segur, H., *Studies in Applied Math.*, 1991, **85,** 129–181.
9. Poincaré, H., *Acta Mathematica*, 1886, **8,** 295–344.
10. Segur, H., Tanveer, S. and Levine, H., *Editors*, "Asymptotics Beyond All Orders," New York: Plenum Press, 1991.
11. Olver, F. W. J., "Asymptotics and Special Functions", New York: Academic Press, 1974.
12. Meyer, R., *SIAM Review*, 1989, **31,** 435–445.
13. Stieltjes, T. J., "*Recherches sur quelques séries semi-convergentes.*" *Ann. Sci. Ecole Norm.*, Paris, 1886, **3,** 201–258. Reprinted in "Oeuvres Complètes," Noordhoff: Groningen, 1918, vol. 2, pp. 2–58.
14. Olver, F. W. J., *SIAM Review*, 1980, **22,** 188–203.
15. Dingle, R. B., "Asymptotic Expansions: their Derivation and Interpretation," New York and London: Academic Press, 1973.
16. Gradshteyn, I. S. and Ryzhik, I. W., "Tables of Integrals, Series and Products," 4th edn., New York: Academic Press, 1965, p. 495.
17. Olver, F. W. J., "On Stokes' phenomenon and converging factors." *In* "Asymptotic and Computational Analysis," R. Wong, *Editor*. New York: Marcel Dekker, 1990, pp. 329–355.
18. Olver, F. W. J., *SIAM J. Math. Anal.*, 1991, **22,** 1475–1489.
19. Jones, D. S., "Uniform asymptotic remainders." *In* "Asymptotic and Computational Analysis," R. Wong, *Editor*, New York: Marcel Dekker, 1990, pp. 241–264.
20. Boyd, W. G. C., *Proc. Roy. Soc. Lond.*, 1990, **A429,** 227–246.
21. Paris, R. B., *J. Comp. Appl. Math.*, 1992, **41,** 117–133.
22. Berry, M. V., *Proc. Roy. Soc. Lond.*, 1990, **A427,** 265–280.
23. McLeod, J. B., *Proc. Roy. Soc. Lond.*, 1992, **A437,** 343–354.
24. Olver, F. W. J., *SIAM J. Math. Anal.*, 1993, **24,** 756–767.
25. Olde Daalhuis, A. B. and Olver, F. W. J., "Exponentially-improved asymptotic solutions of ordinary differential equations II: irregular singularities of rank one." *Proc. Roy. Soc. Lond.*, 1994, **A445,** 39–56.
26. Paris, R. B., *Proc. Roy. Soc. Lond.*, 1992, **A436,** 165–186.
27. Paris, R. B. and Wood, A. D., *J. Comp. Appl. Math.*, 1992, **41,** 135–143.
28. Berry, M. V., *Proc. Roy. Soc. Lond.*, 1991, **A434,** 465–472.
29. Lawless, F., "Exponential asymptotics and spectral theory for optical tunnelling." Ph.D. Thesis, Dublin City University, 1993.
30. Olde Daalhuis, A. B., *IMA Journal of Appl. Math.*, 1992, **49,** 203–216.
31. Berry, M. V. and Howls, C. J., *Proc. Roy. Soc. Lond.*, 1991, **A434,** 657–675.
32. Howls, C. J., *Proc. Roy. Soc. Lond.*, 1992, **A439,** 373–396.