STOKES' PHENOMENON; SMOOTHING A VICTORIAN DISCONTINUITY

par M. V. BERRY

Dedicated to René Thom.

Summary

Small exponentials in asymptotic representations of functions $y(k; \mathrm{X})$ $(k \to \infty)$ can appear and disappear across sets of codimension 1 in the space of variables X. These changes are not discontinuous but happen smoothly and according to a universal law.

1. Problem

My aim* is to present a new result in asymptotics, with a strong connection to some of René Thom's beautiful ideas about singularities. As will become obvious, the treatment is far from rigorous, and the same is true of a more technical version being published elsewhere [1]. But I gain comfort from the conjectured converse of one of Thom's aphorisms: what is non-rigorous might not be insignificant. The work is however insignificant in (at least) one respect, because it deals with exponentially small quantities, which are frequently negligible (and more frequently neglected).

Stokes' phenomenon concerns the behaviour of small exponentials whilst hidden behind large ones. A simple context in which it arises is the approximation of integrals

$$y(k; \mathrm{X}) \equiv \int_{\mathrm{C}} ds\, g(s; \mathrm{X}) \exp\{k\Phi(s, \mathrm{X})\} \tag{1}$$

as $k \to \infty$. Here C is an infinite contour in the complex s plane and Φ is an analytic function of s depending also on variables $\mathrm{X} = (\mathrm{X}_1, \mathrm{X}_2 \ldots)$. Asymptotically, contributions can come from *critical points* (saddles) of Φ, i.e. $s = s_j(\mathrm{X})$ where

$$\partial_s \Phi\{s_j(\mathrm{X}), \mathrm{X}\} = 0. \tag{2}$$

---

* Lecture delivered at the Colloque René Thom, Paris, September 26-October 1 1988.

To isolate these, it is customary to deform C to pass through the different accessible critical points on paths of steepest descent. These paths lie along gradient lines of Re Φ, that is level lines of Im Φ. The (complex) heights of the critical points are

$$\varphi_j(X) \equiv \Phi\{ s_j(X); X \}. \tag{3}$$

Each critical point gives an exponential contribution to y. The dominant contribution has the largest value of Re φ_j; other contributions (subdominant) are exponentially smaller.

As X varies, the steepest-descent contours can change discontinuously in two different ways, illustrated in figures 1 and 2. First, critical *points* s_j can coalesce; this happens on the (complexified) catastrophe set in the X-space [2], and corresponds to large values of y because on the set the critical point is of higher order. Second, critical *values* Im φ_j can coalesce; this happens on the Stokes set in the X-space [3] and corresponds to the appearance or disappearance of a subdominant exponential in a " non-local bifurcation " [2]. The Stokes set is a subset of the saddle-connection set (there may

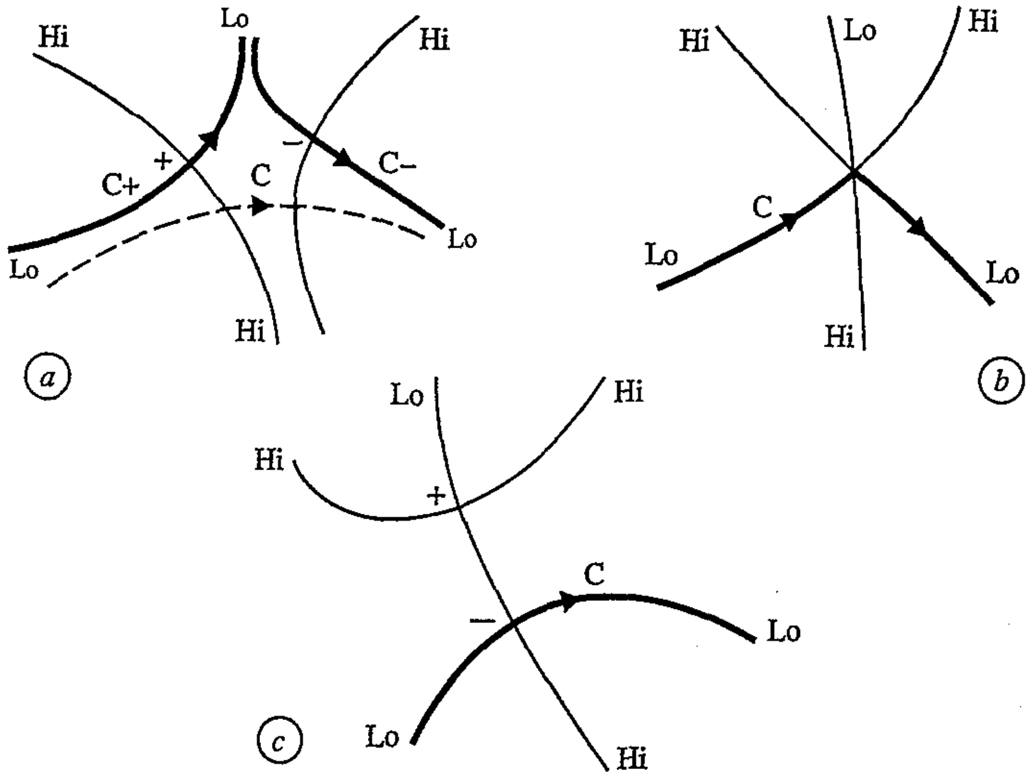

FIG. 1. — Coalescence and separation of saddles in s plane as X varies across the catastrophe set. Hi and Lo denote asymptotic ridges and valleys of Re Φ; light lines are steepest paths, i.e. contours of Im Φ; heavy lines are steepest descent contours through the saddles + and —; the dashed line in *(a)* is a possible defining contour for the integral. The catastrophe occurs at *(b)*. At *(a)*, both saddles contribute; at *(c)*, one contributes.

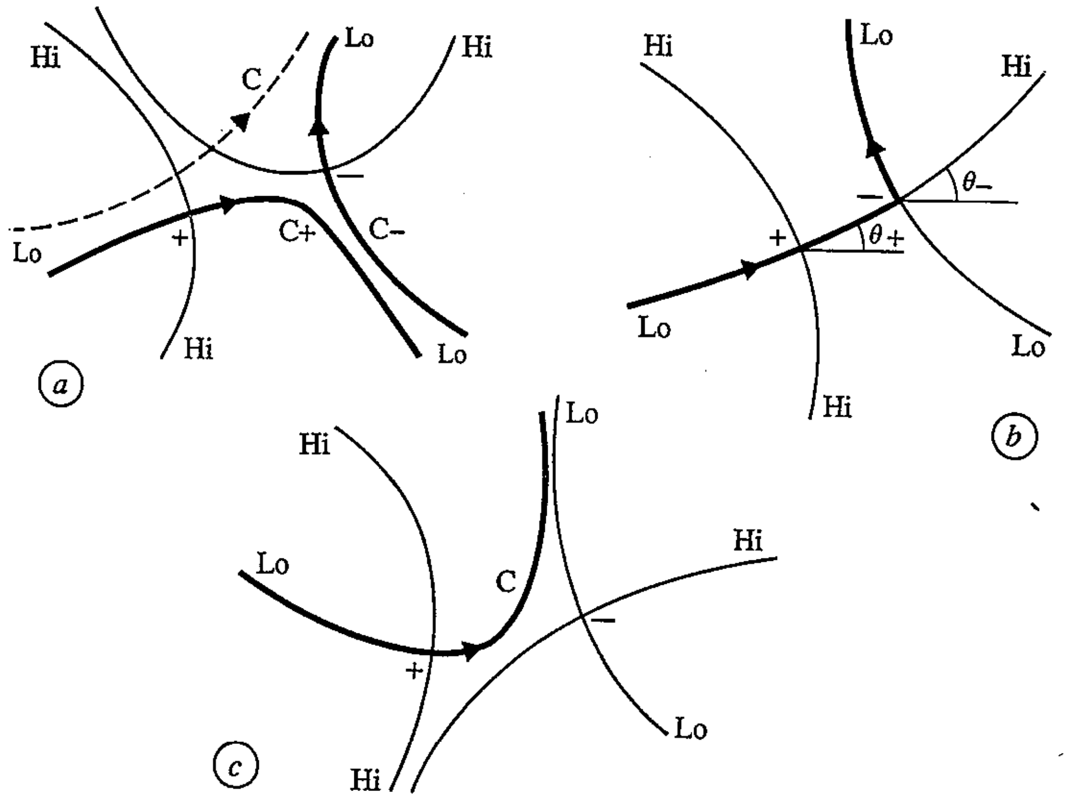

FIG. 2. — Coincidence of heights Im Φ of saddles as X varies across the Stokes set. Notation as in figure 1. *(b)* is on the Stokes set. At *(a)*, both saddles contribute; at *(c)*, one contributes.

be non-contributing connected saddles, through which the deformed C does not pass), which is itself a subset of the Maxwell set, consisting of those X for which any pair of Im φ_j are equal (if the corresponding s_j are distant they need not be connected by a level line). Although the Stokes set has codimension 1 and so is a hypersurface in X, it is commonly called the Stokes *line* because in examples X is often two-dimensional (e.g. the plane of a complex variable $Z = X_1 + iX_2$).

It is the second case with which we are concerned here, because the fact that small exponentials in asymptotic representations can appear and disappear as X varies is the Stokes phenomenon. In the general case we have, to leading order,

$$y(k; X) = M_+(k; X) \exp\{k\Phi_+(X)\} + iS(k; X)\, M_-(k; X) \exp\{k\Phi_-(X)\} + \dots \tag{4}$$

Here + and − denote the dominant exponential and the principal subdominant one (i.e. $\mathrm{Re}\,\varphi_+ > \mathrm{Re}\,\varphi_-$), the prefactors M_+ and M_- are slowly-varying functions of k and X, and ... denotes any further (smaller) exponentials and asymptotic corrections (in higher powers of k^{-1}) to the leading terms. The quantity of principal interest is the *Stokes multiplier* S, whose increase from 0 to 1 describes the switching-on of the small exponential across the Stokes line. S can vary rapidly with k and X. (The factor i is included for later convenience.)

Stokes' opinion was that S varies discontinuously, even though y is continuous. After half a century's reflection on the subject, he wrote [4]

> "... the inferior term enters as it were into a mist, is hidden for a little from view, and comes out with its coefficient changed. The range during which the inferior term remains in a mist decreases indefinitely as the [asymptotic parameter] increases indefinitely."

He had come to this view by analyzing [5] the divergence of the principal asymptotic series which begins with the first term in [4], namely

$$y(k; X) = M_+ \exp\{k\Phi_+\} \sum_{r=0}^{\infty} a_r \quad (a_0 \equiv 1; a_r \propto k^{-r}). \tag{5}$$

The coefficients a_r decrease and then increase. Stokes found that away from the Stokes line the phases of the a_r vary, causing a degree of cancellation which enabled him to perform a crude resummation of the divergent tail of the series. On the Stokes line, however, the a_r all have the same phase and he was unable to resum the series. He concluded that the divergence is incurable, and that after summing down to the smallest a_r the asymptotic expansion specifies y only up to an irremovable vagueness. This vagueness is just sufficient to allow the discontinuous emergence of the small exponential.

Stokes' understanding of his phenomenon was not won easily. On 19 march 1857 he described his discovery in a letter to his fiancée, Mary Robinson [6]:

> "When the cat's away the mice may play. You are the cat and I am the mouse. I have been doing what I guess you won't let me do when we are married, sitting up till 3 o'clock in the morning fighting hard against a mathematical difficulty. Some years ago I attacked an integral of Airy's, and after a severe trial reduced it to a readily calculable form. But there was one difficulty about it which, though I tried till I almost made myself ill, I could not get over, and at last I had to give it up and profess myself unable to master it. I took it up again a few days ago, and after a two or three days' fight, the last of which I sat up till 3, I at last mastered it. I don't say you won't let me work at such things, but you will keep me to more regular hours. A little out of the way now and then does not signify, but there should not be too much of it. It is not the mere sitting up but the hard thinking combined with it..."

He had to fight so hard with his discontinuity because he mistakenly strove to relate it to a superficially similar one he had explored in detail ten years before [7], and which is now commonly attributed to Gibbs, who rediscovered it half a Century later: the ability of Fourier series to represent discontinuous functions, by converging more slowly near discontinuities.

2. Solution

The new result I will present here derives from the fact that it *is* possible to resum the divergent series of a_r beyond the least term, even on the Stokes line, and thereby control the asymptotics of y to an exponential accuracy in k, sufficient to establish the precise variation of S across the line. The variation is not discontinuous but smooth. Moreover the multiplier is *universal* in form, that is the same for all functions in a wide class. I will state the result and list the elements of its derivation; details appear elsewhere [1].

The natural measure of disparity between the dominant and subdominant exponentials is the *singulant* [8]

$$\mathrm{F}(k;\mathrm{X}) \equiv k\{\varphi_+(\mathrm{X}) - \varphi_-(\mathrm{X})\}. \tag{6}$$

On the Stokes line, F is positive real. In terms of the *Stokes variable*

$$\sigma(k;\mathrm{X}) \equiv \operatorname{Im}\mathrm{F}/(2\operatorname{Re}\mathrm{F})^{1/2} \tag{7}$$

describing the crossing of the Stokes line, the Stokes multiplier is

$$\mathrm{S}(\sigma) = \mathrm{S}_- + \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\sigma} dt\,\exp(-t^2) \tag{8}$$

where S_- is the value of the multiplier below the line (i.e. for $\operatorname{Im}\mathrm{F} \leqslant 0$).

This result is surprising, because it shows that hidden in the asymptotics of a huge variety of functions (including the special functions of classical analysis—Bessel, hypergeometric, etc.—and the diffraction catastrophes of optics [9]) is the humble error function. It is revealed by subtracting from y the dominant series (5), summed to its least term, that is

$$\lim_{k\to\infty} -i\mathrm{M}_-^{-1}\{y\exp(-k\varphi_-) - \mathrm{M}_+\exp(\mathrm{F})\sum_{r=0}^{r^*} a_r\} = \mathrm{S}(\sigma) \tag{9}$$

where $|a_{r^*}| < |a_r|$ $(r \neq r^*)$.

Before outlining the derivation I should remark that it is impossible to study Stokes' phenomenon within the framework of Poincaré's definition of an asymptotic expansion. This states that a_r are asymptotic coefficients for y if

$$\lim_{k\to\infty} k^m\{\mathrm{M}_+^{-1}\exp(-k\varphi_+)\,y - \sum_{r=0}^{m} a_r\} = 0(1). \tag{10}$$

It is inadequate because it captures the asymptotics of y only to power-law accuracy, whereas understanding Stokes' multiplier requires exponential accuracy.

3. Derivation

The derivation of (9) is based on an interpretation of the divergent dominant series (5). It is not necessary to include the subdominant exponential because it will be born out of the resummed tail of (5). Ecalle [13] has coined the term " resurgence " to describe this phenomenon. Resurgence appears remarkable but is in fact inevitable because the existence of the small exponential is the cause of the divergence of the dominant series—convergent series in powers of k^{-1} would not be able to represent an exponential. We employ *Borel summation* of (5), starting with the least term r^*. This method has been extensively developed by Dingle [8] and applied to the approximation of y on the Stokes line itself; here it is applied *across* the line.

Underlying the universality of the multiplier (8) is a universality in the form of the *late terms* a_r ($r \gg 1$), that is in the asymptotics of the asymptotics:

$$a_r \underset{r\to\infty}{\sim} \frac{M_-}{2\pi M_+} \frac{(r-\beta)!}{F^{r-\beta}}. \tag{11}$$

Dingle derives (11) (and corrections to it) for integrals of the form (1) (where + and − correspond to stationary points, and where $\beta = 1$), for integrals with finite contours (where + comes from an end point and − from a stationary point, and where $\beta = 1/2$), and for second-order linear differential equations with variable coefficients (where + and − describe waves running in opposite directions, and where $\beta = 1$). To show how the universality emerges, I give in the appendix a derivation of the late terms for the first of these cases.

From (11), the least term has

$$r^* \approx |F|. \tag{12}$$

(The precise value is immaterial because changing r^* by one contributes a correction of order $k^{-1/2}$ to $S(\sigma)$, which is invisible in the limit (9).)

Borel summation gives an integral representation for the tail of the series, that is for the sum of terms $r > r^*$. A crucial simplification is that truncation near the least term (i.e. r^* given by (12)) gives a Borel integral with a stationary point coinciding with a pole, whose approximation (that is, the asymptotics of the asymptotics of the asymptotics) is quite easy and yields our results (8) and (9).

Numerical tests of (9) [1] show nicely how the error function (8) emerges, and the robustness of the results under changes in the truncation r^*, even when the asymptotic parameter as measured by $|F|$ is not particularly large (e.g. $|F| = 5$).

4. Stokes and Airy

It is instructive to examine the numerical calculation performed by Stokes himself [5] to establish the reality of his phenomenon. He was studying the integral

$$\mathrm{A}i(z) \equiv \frac{1}{2\pi}\int_{-\infty}^{\infty} ds \exp\{i(s^3/3 + zs)\} \tag{13}$$
$$(z = X_1 + iX_2)$$

introduced by Airy [10] in 1838 to describe difraction near a caustic (e.g. the rainbow). In optics one needs the values of $\mathrm{A}i(z)$ for z real, i.e. $X_2 = 0$. For negative X_1 the function oscillates in characteristic interference fringes (describing, for example, supernumerary rainbows inside the main arc). But Airy was unable to compute these fringes numerically because the only technique available to him (representing A*i* by a convergent series) was limited to too small values of $|X_1|$.

Stokes solved this problem [11] by inventing what we now call the WKB method, applied to the differential equation satisfied by A*i*, to calculate asymptotic (divergent) series enabling A*i* to be computed to high accuracy for large $|X_1|$. He also, almost in passing, invented what we now call the method of stationary phase.

The integral (13) has two stationary points, at $s = \pm(-z)^{1/2}$. When z is real and $X_1 \gg 0$ only one of these contributes to the integral, which is exponentially small. When z is real and $X_1 \ll 0$ both stationary points contribute and A*i* oscillates (figure 3). This was the source of the difficulty which occupied Stokes for so long. How could one function have two asymptotic expansions (for $X_1 > 0$ and $X_1 < 0$)? The resolution of course lay in studying A*i* for complex z. Somewhere between the positive and negative real axes, a second exponential must be born. This happens near the Stokes lines, which for A*i* lie at $\arg(z) = 120°$ and $240°$.

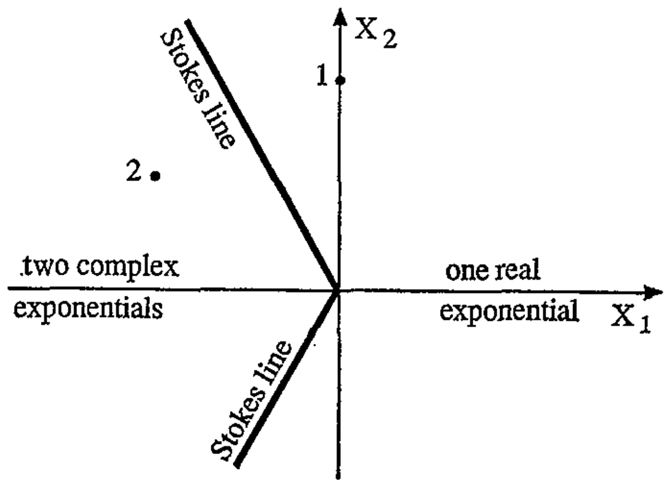

FIG. 3. — Points 1 and 2, on opposite sides of a Stokes line at which Stokes computed Airy's function A*i*(z)

To test his theory, Stokes computed A*i* at two points (labelled 1 and 2 on figure 3) on opposite sides of the 120° line, with $\arg(z) = 90°$ and $150°$ and $|z| = (72)^{1/3} = 4.160...$ For these points, the singulant modulus is $|F| = (128)^{1/2} = 11.31...$ He computed A*i* "exactly" (from the convergent series) and from the divergent series for the dominant exponential, taken to its least term. The results [5] are reproduced in table I. At point 1 this series approximates A*i* to one part in 10^7. At point 2 it is accurate to only one part in 10^4—that is 10^3 times worse. But the accuracy is restored at point 2 by including just the leading term of the subdominant exponential, thereby establishing the reality of the Stokes phenomenon. Several authors (e.g. [12]) have rediscovered the dramatically increased accuracy that results when exponentially small terms are correctly added to optimally truncated dominant series.

TABLE I. — Stokes' computations of the Airy integral (13) (multiplied by $2\sqrt{\pi}3^{1/6}$) for $|z| = (72)^{1/3}$ and $\arg(z) = 90°$ (point 1) and $150°$ (point 2)

Value of integral	Point 1	Point 2
Exact	$-14.98520 + i\,43.81047$	$-45.44882 - i\,8.92867$
Dominant series	$-14.98520 + i\,43.81046$	$-45.43360 - i\,8.92767$
Subdominant		$-\ \ 0.01524 - i\,0.00100$
Total	$-14.98520 + i\,43.81046$	$-45.44884 - i\,8.92867$

Stokes' computations were consistent with his opinion that the multiplier changes discontinuously. He missed the fact that S varies smoothly (cf. (8)) because his points 1 and 2 are too far from the Stokes line. At 1, the Stokes variable is $\sigma = -2$, and $S(\sigma) = .005$, so that the birth of the second exponential has hardly begun. At 2, $\sigma = +2$, and $S(\sigma) = .995$, so that the birth is virtually complete.

In a sense the result reported here completes a story begun by Airy and Stokes. Airy realized that the singularity at a caustic is an artefact of ray theory which would be smoothed away by properly taking diffraction into account. His function $Ai(z)$ accomplishes this smoothing in the generic case, which we now know as the fold diffraction catastrophe [2]. For integrals, the discontinuity thus smoothed is the one illustrated in figure 1. Stokes discovered that in the complex z plane $Ai(z)$ itself has discontinuities in its asymptotic representation, of the other kind as illustrated in figure 2. He did not however find the appropriate smoothing. That is accomplished by our result (8) and (9), which shows " the error function in the Airy function ". (The same smoothing occurs across the Stokes line for the error function, which is therefore contained in its own asymptotic approximation—" the error function in the error function ".)

5. Discussion

I envisage several applications of this work, beyond the purely numerical. In wave optics the Stokes set may be observable if there are at least two real variables X. This is not the case for the fold caustic because the Stokes lines in $z = X_1 + iX_2$ are complex and in diffraction we usually have z real. But for the higher catastrophes the Stokes set can be real. Wright [3] has calculated it for the cusp diffraction catastrophes, and work is in progress on the higher singularities. Observation of the Stokes phenomenon would be difficult (if possible at all) because it involves exponentially weak complex rays masked by intense real rays. (The situation with caustics—singularities of the other sort—is quite different: these are sets of high intensity, dominating wave fields.) Other applications are to the birth of weak reflected waves in smooth refractive index gradients, and the generation of weak nonadiabatic jumps in slowly-varied parametric oscillators.

There must be limits to the universality of our smoothing (8) and (9), reflecting limits in the universality of the asymptotics of the asymptotics (11). Presumably the breakdown of universality occurs when Stokes lines coalesce or cross as more variables X are altered. There ought to be a classification of the ways in which this can happen stably, and of the associated smoothings, analogous to the classification of catastrophes and their associated diffraction patterns.

Acknowledgements

I am most grateful for the generous hospitality of the Physical Research Laboratory, Ahmedabad, during the writing of this paper, and to the organizers of the Colloque René Thom for inviting the lecture on which it was based.

Appendix

This is the derivation (following [8]) of the "asymptotics of the asymptotics" giving the late terms (11) of the integral (1). First the prefactors $M_\pm$ must be found. The lowest-order approximation to (1) from a dominant saddle s_0 is

$$y \approx \left[\frac{2\pi}{k\,|\,\Phi_0''\,|}\right]^{1/2} g_0 \exp(k\Phi_0 + i\theta_0) \tag{A1}$$

where dashes denote s-derivatives, the subscripts quantities evaluated at s_0, and θ_0 is the direction in which the deformed contour of steepest descent departs from s_0. Referring to figure 2b, let s_0 now be the dominant saddle s_+ and (without loss of generality) choose the sense of C towards the principal subdominant saddle s_-. Then we can take

$$M_\pm = \left[\frac{2\pi}{k\,|\,\Phi_\pm''\,|}\right]^{1/2} g_\pm \exp(i\theta_\pm) \tag{A2}$$

where (for real singulant F (i.e. on the Stokes line) θ_- is the direction in which the level curve $\operatorname{Im} F = 0$ through s_+ emerges from s_- (where it is a path of steepest ascent)).

Now change variables in (1) from s to w, defined near s_+ by

$$k\Phi(s) = k\varphi_+ - w^2/2. \tag{A3}$$

Thus w is real and increases from zero along the steepest path from s_+ to s_-. Expansion in powers of w gives the formally exact expression

$$y = \exp(k\varphi_+) \sum_{r=0}^{\infty} \frac{2^{r+1/2}\,\Gamma(r+1/2)}{(2r)!} A_{2r} \equiv \sum_{r=0}^{\infty} y_r \tag{A4}$$

in which

$$\begin{aligned} A_{2r} &= \frac{d^{2r}}{dw^{2r}}\left\{g[s(w)]\,\frac{ds(w)}{dw}\right\} \\ &= \frac{(2r)!}{2\pi i}\oint \frac{dw}{w^{2r+1}}\, g\, \frac{ds}{dw} \end{aligned} \tag{A5}$$

where the contour is a small loop around $w = 0$.

The late terms $r \gg 1$ are found by expanding the contour until it meets the nearest singularity. This is the w corresponding to s_-, namely

$$w = [2k(\varphi_+ - \varphi_-)]^{1/2} = (2F)^{1/2}. \tag{A6}$$

To find the form of the singularity, expand (A3) about s_- to get

$$(s - s_-)^2 k\Phi''_- = F - w^2/2. \tag{A7}$$

Inversion and differentiation now give

$$\frac{ds}{dw} \underset{s \to s_-}{\sim} \frac{\exp(i\theta_-)\,(2F)^{1/4}}{\{2\Phi''_-[(2F)^{1/2} - w]\}^{1/2}} \tag{A8}$$

(the phase is determined by the direction of the level line from s_-—cf. figure 2b).

The leading term of the integral (A5) is given by the integral along the sides of the cut emerging along the positive w axis from $w = (2F)^{1/2}$. Noting that the phase of the radical in (A8) on the upper lip of this cut is $-\pi/2$, we obtain

$$A_{2r} \approx \frac{(2r)!\,g_-\exp(i\theta_-)\,(2F)^{1/4}}{\pi(2\Phi''_-)^{1/2}} \int_0^\infty \frac{dx}{[(2F)^{1/2} + x]^{2r+1}\,x^{1/2}} \tag{A9}$$

where $x = w - (2F)^{1/2}$. For large r the integrand decays exponentially away from $x = 0$, giving

$$A_{2r} \underset{r \to \infty}{\sim} \frac{(2r)!\,g_-\exp(i\theta_-)}{\pi^{1/2}(2F)^r\,[(2r+1)\,2\Phi''_-]^{1/2}}. \tag{A10}$$

Substitution into (A4) and use of $\Gamma(r + 1/2)/(r + 1/2)^{1/2} \sim (r - 1)!$ gives

$$\begin{aligned} y_r &\underset{r \to \infty}{\sim} g_-\exp(i\theta_-)\left(\frac{2\pi}{|\Phi''_-|}\right)^{1/2} \exp(k\varphi_+)\,\frac{(r-1)!}{2\pi F^r} \\ &= M_-\exp(k\varphi_+)\,\frac{(r-1)!}{2\pi F^r} \end{aligned} \tag{A11}$$

which is the same as (11) with $\beta = 1$.

REFERENCES

[1] M. V. Berry, Uniform asymptotic smoothing of Stokes's discontinuities, *Proc. Roy. Soc. Lond.*, **A422** (1989), 7-21.

[2] T. Poston and I. N. Stewart, *Catastrophe theory and its applications*, London, 1978.

[3] F. J. Wright, The Stokes set of the cusp diffraction catastrophe, *J. Phys.*, **A13** (1980), 2913-2928.

[4] G. G. Stokes, On the discontinuity of arbitrary constants that appear as multipliers of semi-convergent series, *Acta. Math.*, **26** (1902), 393-397, reprinted in *Mathematical and Physical Papers by the late Sir George Gabriel Stokes*, Cambridge University Press, 1905, vol. V, p. 283-287.

[5] G. G. Stokes, On the discontinuity of arbitrary constants which appear in divergent developments, *Trans. Camb. Phil. Soc.*, **10** (1864), 106-128, reprinted in *Mathematical and Physical Papers*... (ref. [4]), vol. IV, p. 77-109.

[6] J. LARMOR (ed.), *Sir George Gabriel Stokes: Memoir and Scientific Correspondence* (Cambridge University Press, 1907), vol. 1, p. 62.

[7] G. G. STOKES, On the critical values of the sums of periodic series, *Trans. Camb. Phil. Soc.*, **8** (1847), 533-610, reprinted in *Mathematical and Physical Papers*... (ref. [4]), vol. I, p. 236-313.

[8] R. B. DINGLE, *Asymptotic Expansions: their Derivation and Interpretation*, New York and London, Academic Press, 1973.

[9] M. V. BERRY and C. UPSTILL, Catastrophe optics: morphologies of caustics and their diffraction patterns, *Progress in Optics*, **18** (1980), 257-346.

[10] G. B. AIRY, On the intensity of light in the neighbourhood of a caustic, *Trans. Camb. Phil. Soc.*, **6** (1838), 379-403.

[11] G. G. STOKES, On the numerical calculation of a class of definite integrals and infinite series, *Trans. Camb. Phil. Soc.*, **9** (1847), 379-407, reprinted in *Mathematical and Physical Papers*... (ref. [4]), vol. II, p. 329-357.

[12] R. BALIAN, G. PARISI and A. VOROS, Discrepancies from asymptotic series and their relation to complex classical trajectories, *Phys. Rev. Lett.*, **41** (1978), 1141-1144.

[13] J. ECALLE, *Les fonctions résurgentes* (3 vol.), Publ. Math. Université de Paris-Sud, 1981, and *Cinq applications des fonctions résurgentes*, preprint 84T62, Orsay, 1984.

H. H. Wills Physics Laboratory
Tyndall Avenue
Bristol BS8 1TL
UK

Manuscrit reçu le 4 janvier 1989.