The Stokes set of the cusp diffraction catastrophe

1980 J. Phys. A: Math. Gen. 13 2913

(http://iopscience.iop.org/0305-4470/13/9/018)

J. Phys. A: Math. Gen. **13** (1980) 2913–2928. Printed in Great Britain

The Stokes set of the cusp diffraction catastrophe

F J Wright†

H H Wills Physics Laboratory, Bristol University, Royal Fort, Tyndall Avenue, Bristol, BS8 1TL, UK

Received 12 February 1979, in final form 26 February 1980

Abstract. Stokes and anti-Stokes lines are familiar in the asymptotic approximation of functions of a complex variable. We generalise this notion and define the Stokes and anti-Stokes sets of a complex function of many (possibly complex) variables, defined by a diffraction-type integral. They are subsets of the Maxwell set of catastrophe theory, extended to complex variables. On the Stokes set the number of complex stationary points contributing to the integral changes by one, whereas on the caustic the number of real stationary points changes by two. Knowledge of the location of the Stokes set is essential to perform a full stationary phase analysis of a diffraction integral, and it imposes a constraint upon the positions of wavefront dislocations. For the canonical cusp diffraction catastrophe we find the explicit equation of the Stokes set, which is a broadened mirror image of the cusp caustic.

1. Introduction

Catastrophe theory (Thom 1975, Poston and Stewart 1978) is concerned with the forms of real functions. Consider a family, parametrised by real control variables $\boldsymbol{c}$, of smooth real functions $\phi(\boldsymbol{s};\boldsymbol{c})$ of the real state variables $\boldsymbol{s}$. The form of a smooth function is largely determined by its stationary points, which for a member of our family of functions are points $\boldsymbol{s}$ at which

$$\nabla_{\boldsymbol{s}}\phi(\boldsymbol{s};\boldsymbol{c})=0.$$

The family of functions divides into subfamilies having different forms, which implies that the control space is divided into regions separated by a subset of co-dimension one (i.e. dimension one less than the space), called by Thom (1975, p 57) the catastrophe set. Thom distinguishes two types of catastrophe point: conflict points and bifurcation points. At a conflict point $\phi(\boldsymbol{s};\boldsymbol{c})$ has two or more absolute minima. At a bifurcation point two or more stationary points coalesce. On crossing the conflict set C the absolute minimum of $\phi(\boldsymbol{s};\boldsymbol{c})$ changes from one stationary point to another, while on crossing the bifurcation set B the number of real stationary points changes by two, which is usually much more significant physically.

The families of functions fall into a discrete set of classes, each of which constitutes one catastrophe. Every member of a class has the same form, which is represented by its simplest member—the canonical catastrophe polynomial or *normal form*, such as equation (2) below for the cusp catastrophe.

† Now at the Department of Applied Mathematics, Queen Mary College, University of London, Mile End Road, London, E1 4NS, UK.

0305-4470/80/092913+16$01.50

The form of a function is well displayed by its level lines

$$\phi(\boldsymbol{s};\boldsymbol{c}) = \text{constant},$$

in state space, for particular values of $\boldsymbol{c}$. This is particularly convenient when $\boldsymbol{s}$ is two-dimensional, and for ease of description we shall often assume this to be the case. The connectivity of the level lines changes on crossing B, but not on crossing C, because the level lines near an extremum are compact.

Arnol'd (1974) generalised the conflict set by defining the *Maxwell set* M to be the control points at which stationary *values* coincide. Poston and Stewart (1978, p 239) refer to this as the *non-local bifurcation set*, as against the *local bifurcation set* B at which stationary *points* coincide. M is important if the *global* form of a function is relevant. The connectivity of the level lines of the function changes not only on B, but also on a subset of M called by Poston and Stewart the *saddle-connection set*. This is the set of control points at which stationary points are joined by a level line (which clearly excludes extrema). The significance of the saddle-connection set shows up clearly when the level lines represent, for example, the stream lines of a fluid flow, sucn as that analysed by Berry and Mackley (1977).

We shall generalise the non-local bifurcation set further, by moving outside the scope of catastrophe theory proper and allowing $\boldsymbol{s}$, and possibly also $\boldsymbol{c}$, to become complex. Such complexification is essential in the application of catastrophe theory to diffraction. We can describe the form of the complex function $\phi(\boldsymbol{s};\boldsymbol{c})$ in terms of its real and imaginary parts ϕ_R and ϕ_I respectively. For cuspoid catastrophes $\boldsymbol{s}$ is a single variable s, which we have allowed to become complex. Since ϕ is an analytic function of s, ϕ_R and ϕ_I are not independent, but related by the Cauchy–Riemann equations. This means that the level lines of ϕ_R and ϕ_I in the complex s plane form an orthogonal net, and that the stationary points of ϕ can only be saddles. Hence Thom's idea of conflict set is now quite inappropriate, but we can consider the Maxwell and saddle-connection sets of ϕ_R and ϕ_I, as real functions of s_R and s_I. The Stokes and anti-Stokes sets, which are the subject of this paper, are subsets of the saddle-connection set of ϕ_R and the Maxwell set of ϕ_I respectively.

2. Diffraction catastrophes

In diffraction theory the functions $\phi(\boldsymbol{s};\boldsymbol{c})$ are source–receiver distances or path lengths. The state variables $\boldsymbol{s}$ parametrise the source points, and the control variables $\boldsymbol{c}$ parametrise the receiver or field points. A stationary point of ϕ corresponds to a geometrical ray. If two or more geometrical rays coalesce, a caustic occurs, which in the geometrical model produces infinite intensity; i.e. field points $\boldsymbol{c}$ at which two or more points, where $\phi(\boldsymbol{s};\boldsymbol{c})$ is stationary with respect to $\boldsymbol{s}$, coalesce lie on a caustic. (For further details see Berry (1976), Poston and Stewart (1978, ch 12) and the review by Berry and Upstill (1980).) We see that the caustic in the observation space is precisely the bifurcation set B. The Maxwell set M does not manifest itself in applications to diffraction.

For waves of finite wavelength, the geometrical caustic is clothed with a diffraction pattern, now called, following Trinkaus and Drepper (1977), a diffraction catastrophe. This is described by the set of canonical diffraction integrals

$$\psi(\boldsymbol{c}) = (2\pi)^{-n/2} \int_{-\infty}^{\infty} \mathrm{d}^n\boldsymbol{s}\, \exp(\mathrm{i}\phi(\boldsymbol{s};\boldsymbol{c})), \qquad \boldsymbol{s} \in \mathbb{R}^n \tag{1}$$

generated by the normal form $\phi(s; c)$ for each of the catastrophes (Berry 1976). This set of canonical diffraction integrals forms the set of standard functions with which to analyse any diffraction catastrophe, by using them as the comparison integrals in a uniform approximation (Berry 1976). Moreover, for short wavelengths and close to the caustic the appropriate canonical integral on its own (apart from various wavelength scaling factors) provides a very good approximation to the diffraction catastrophe (see Berry, Nye and Wright (1979), henceforth referred to as BNW). The status of these canonical integrals is that they play the same role in the analysis of caustic wavefields that complex exponentials play in the elementary analysis of plane waves: they are important mathematical tools, and as such need to be thoroughly understood.

Catastrophes, and hence diffraction catastrophes, have a hierarchical structure: every catastrophe (except the fold) contains simpler *subordinate* catastrophes (e.g. Arnol'd 1973, Connor 1976). Therefore the way to study a catastrophe is to approach it via its subordinate components. The simplest catastrophe is the *fold*, and its associated canonical diffraction integral is essentially the well-known Airy integral function (Abramowitz and Stegun 1964) introduced by Airy (1838) to study the rainbow, probably the most familiar caustic.

The second member of the hierarchy is the cusp catastrophe, whose normal form is

$$\phi(\xi; x, y) = \tfrac{1}{4}\xi^4 + \tfrac{1}{2}x\xi^2 + y\xi. \tag{2}$$

The canonical cusp diffraction integral, our principal object of study, is

$$C(x, y) = (2\pi)^{-1/2} \int_{-\infty}^{\infty} \mathrm{d}\xi \exp(\mathrm{i}\phi(\xi; x, y)). \tag{3}$$

It was introduced and studied initially by Pearcey (1946) as a generalisation of the Airy function, and has subsequently been studied by, among others, Trinkaus and Drepper (1977) and BNW. The fold and cusp catastrophes are particularly important because they are subordinate to *every* higher catastrophe. They were used, for example, in the analysis by BNW of the elliptic umbilic diffraction catastrophe, which is represented by a two-dimensional integral.

The diffraction integrals arise physically as integrals over the real state space, but once defined they may be regarded mathematically as contour integrals over a complex state space. This is the step which takes us outside the scope of catastrophe theory proper, and allows us to consider complex state variables, which are now just dummy integration variables. However, for physical diffraction patterns in real space only real control variables occur directly, so in this paper control variables will always be real unless otherwise stated.

The integrands of diffraction integrals are highly oscillatory, so that their behaviour is most usefully described in terms of their amplitude $|\exp(\mathrm{i}\phi)| \equiv \exp(-\phi_\mathrm{I})$ and phase ϕ_R as functions of the complex state variables at each control point. The singular sets in the control space of ϕ—the bifurcation set or caustic, and the Maxwell sets of ϕ_R and ϕ_I—determine the *topography* of the integrand.

Using the canonical integrals, and understanding their structure, is greatly facilitated by a simple tractable approximation, the stationary phase approximation. This has the great merit of being closely related to the physical interference effects which produce the diffraction patterns in nature. We shall see that a knowledge of the Stokes set is essential to perform a full stationary phase analysis of a diffraction integral. But first we pause to consider the other main feature of diffraction patterns, which are complementary to caustics.

3. Wavefront dislocations

An example of the use of stationary phase approximations is in finding wavefront dislocations, which constitute the 'edges' of wavefronts (Nye and Berry 1974, Wright 1977, 1979). When using a complex wavefunction, it is convenient to define wavefronts as surfaces of constant phase, modulo 2π, of the wavefunction. Then a wavefront can end only where the phase is indeterminate, and since the wavefunction must be differentiable, the amplitude (or modulus) must go to zero at least linearly at this point. Hence a *necessary* condition for a wavefront to end is that the amplitude be zero; it is not sufficient because it includes, for example, the coincident edges of two wavefronts, which correspond to the birth or death of a pair of wavefront dislocations (in which case the amplitude goes to zero quadratically). We can, however, *define* a wavefront dislocation to be the set of points at which the wavefunction is zero and consider the birth/death event as a degenerate case (in fact, a dislocation of strength zero). Adopting this philosophy allows us to find dislocations by consideration of the wave amplitude alone, and then having found them to analyse their detailed structure by considering the phase.

Wavefront dislocations are the absolute global intensity minima of a diffraction pattern. They are generally sets of codimension two, e.g. lines in three dimensions, and form a conceptual framework upon which to hang the diffraction pattern. Describing a complicated three-dimensional wavefield is not easy, but a statement of the configuration of the dislocation lines (global intensity minima) and of the caustic (approximate local intensity maxima) provides a very good basis for such a description. This approach is applied to the three-dimensional elliptic umbilic diffraction catastrophe by BNW.

The diffraction catastrophe wavefunctions are defined by complex integrals, so to find the wavefront dislocations we have to find the zeros of these complex integrals. There are no general exact methods for doing this, so we must resort either to numerical solutions or analytical approximations such as stationary phase. In the latter case we discover that the Stokes set can impose a constraint upon the location of the wavefront dislocations.

4. Stationary phase approximation

We concentrate on the canonical cusp diffraction integral (3), which is a function of two real variables, defined by a single complex contour integral. It illustrates the general principles without getting too complicated, and has the great merit that it is the highest diffraction catastrophe for which *explicit* equations can be found for not only the bifurcation set, but also the Stokes set.

We consider the problem of approximating $C(x, y)$ by the method of stationary phase (Dingle 1973, ch IX, Born and Wolf 1975, appendix III, Budden 1961, ch 15). This method assumes that the main contributions to the integral come from the neighbourhoods of points where the phase

$$\phi(\xi; x, y) = \tfrac{1}{4}\xi^4 + \tfrac{1}{2}x\xi^2 + y\xi$$

is stationary with respect to ξ, i.e. points ξ satisfying

$$\mathrm{d}\phi/\mathrm{d}\xi = \xi^3 + x\xi + y = 0. \tag{4}$$

When the stationary points are all real, the integration contour along the real ξ axis passes through them all.

In the stationary phase method there is normally some parameter k multiplying the function ϕ to give the phase, which is taken to be large, producing an asymptotic series in k. For large k the integrand oscillates very rapidly, thereby cancelling itself out, except near stationary points of ϕ. Therefore the neighbourhoods of the stationary points may be assumed to make independent contributions to the integral, as long as they remain isolated. In wave theory k is the wavenumber ($2\pi/\lambda$, where λ is the wavelength), and in the short-wavelength limit, where caustics dominate the wavefield, k is large. However, k has been scaled out of the canonical integrals (1), so that large k is equivalent to large separation of stationary points, except on the caustic B. The neighbourhood of the caustic must be handled by uniform approximation (Chester *et al* 1957, Berry 1976), which we shall not consider here, so we assume that the stationary points are always isolated.

The caustic of $C(x, y)$ is given by

$$d\phi/d\xi = d^2\phi/d\xi^2 = 0$$

to be the cusped semi-cubical parabola

$$27y^2 + 4x^3 = 0$$

shown in figure 1. On the caustic the stationary points of ϕ are all real, and two coincide. Standard theory of cubics (e.g. Ferrar 1958) applied to equation (4) shows that 'inside' the caustic, where

$$27y^2 + 4x^3 < 0,$$

the stationary points are all real and different, but 'outside' the caustic, where

$$27y^2 + 4x^3 > 0,$$

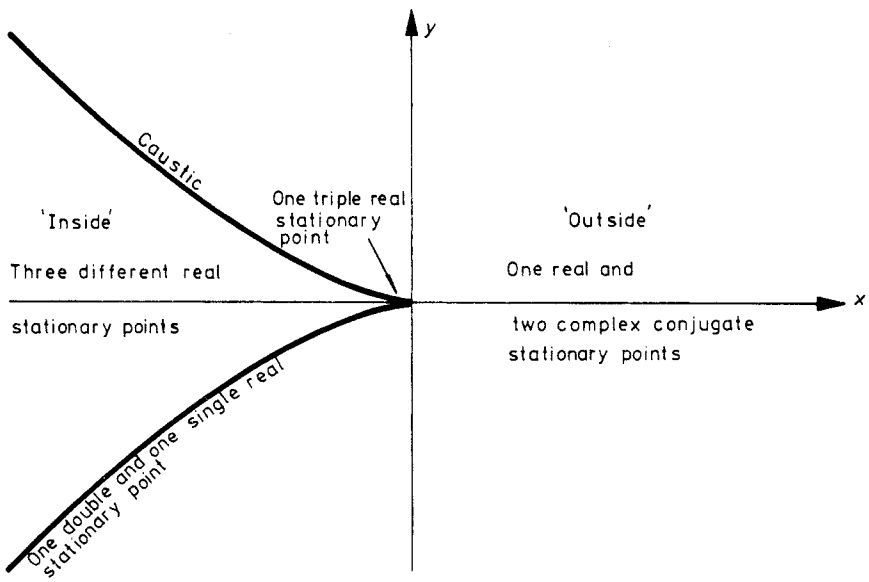

Figure 1. The three regions of the (x, y) plane, in which the roots of the stationary phase equation (4) behave differently.

only one stationary point is real, the other two forming a complex conjugate pair (figure 1).

Inside the caustic the integration contour of (3) along the real ξ axis passes through all the stationary points. We simply add all their individual contributions to find the approximation to $C(x, y)$, as in appendix C of BNW. The three real stationary phase points correspond to three real physical rays (Berry and Upstill 1980).

Outside the caustic we have a problem, because only one of the stationary points lies on the integration contour along the real ξ axis. The simplest approximation is to ignore the other two. This implies that there are no wavefront dislocations, i.e. zeros of $C(x, y)$, outside the caustic, since there is nothing which can cancel the single stationary phase term. From exact computations of $C(x, y)$ (Pearcey 1946, BNW, appendix C) this is known not to be the case—the complex stationary points must be considered in order to retain the full structure of the integral. But which, if either, should be included?

It is at this point that we are forced to regard (3) as a contour integral in the complex ξ plane and employ the method of steepest descent (Dennery and Krzywicki 1967, ch 1, § 31). For some fixed (x, y),

$$|\exp(\mathrm{i}\phi(\xi; x, y))| \equiv \exp(-\phi_{\mathrm{I}}(\xi; x, y))$$

gives the magnitude of the integrand as a function of $(\xi_{\mathrm{R}}, \xi_{\mathrm{I}})$, and its graph is a two-dimensional surface. The stationary points of ϕ are saddle points of this surface; extrema occur only at infinity. The topography of the surface is conveniently summarised by the level lines through its saddles, which are the level lines of ϕ_{I} through its stationary points. (Remember that our exponent is $\mathrm{i}\phi$, not just ϕ).

Since we are only considering real x and y, we have

$$\phi(\xi^*; x, y) = \phi^*(\xi; x, y). \tag{5}$$

Therefore ϕ_{I} is antisymmetric in ξ_{I}, and so the real ξ axis is always a level line of ϕ_{I}, which always passes through a saddle since one is always real. Since ϕ is an analytic function of ξ, the level lines of ϕ_{R} and ϕ_{I} must intersect at 90° at a single saddle, 60° at a double saddle, etc (e.g. Dennery and Krzywicki 1967). As $|\xi| \to \infty$, $\phi(\xi; x, y) \sim \frac{1}{4}\xi^4$ independently of x and y. Putting $\xi = R\exp(\mathrm{i}\theta)$ gives $\phi_{\mathrm{I}} \sim \frac{1}{4}R^4 \sin(4\theta)$, so that asymptotically $|\exp(\mathrm{i}\phi)|$ always has four radial hills H and four radial flat-bottomed valleys V connected by steep cliffs in which all level lines ultimately merge. Using these facts, plus the positions of the saddles, the level lines through the saddles of $|\exp(\mathrm{i}\phi)|$ can be sketched, as in figures 2–4.

For convergence of the integral, the integration contour must end somewhere on the asymptotic zeros of the integrand where $\phi_{\mathrm{I}} \to +\infty$ (i.e. in the valleys V) and avoid the asymptotic singularities where $\phi_{\mathrm{I}} \to -\infty$ (the hills H). A trivial modification of Jordan's lemma (e.g. Dennery and Krzywicki 1967) shows that the contour need not actually end at $\xi = \pm\infty$ on the real axis, but can end at $\xi = \pm\infty\exp(\mathrm{i}\theta)$, where $0 \leqslant \theta \leqslant \frac{1}{4}\pi$.

Subject to these constraints, we distort the integration contour from its original position along the real ξ axis until it lies only along lines of steepest descent from saddles of $|\exp(\mathrm{i}\phi)|$. This may require an excursion into a previously unexplored valley, as in figure 2, and it need not use all the saddles. Once this distortion is complete, the contour is divided into subcontours, each running from a point where the integrand is zero, up to a saddle and back down to a zero (e.g. V_3–S_1–V_2 and V_2–S_2–V_1 in figure 2(*b*)). (See Dingle (1973, ch IX).) All the zeros of this integrand are asymptotic, so both ends of

every subcontour must extend to infinity, although on the figures the subcontours are shown schematically linking together at a finite distance along the valley. The largest value of the integrand along each subcontour occurs at the saddle, and hence the main contribution to the integral comes from its neighbourhood. The contributions from different subcontours are obviously independent and additive, and this is the justification for the independence of real saddle contributions asserted earlier.

The level lines of ϕ_R and ϕ_I form an orthogonal net, and since the level lines of ϕ_I are level lines of $|\exp(i\phi)|$, the level lines of ϕ_R are the lines of steepest descent. Our integration contour runs along level lines of ϕ_R, and, if the topology of these level lines changes, as it does on the saddle-connection set of ϕ_R, then the topology of the integration contour may change. The integration contour only runs along a small number of the level lines of ϕ_R through its saddles, so generally the integration contour changes form only on a subset of the saddle-connection set, which we call the *Stokes set* (for a reason to be explained later). The Maxwell sets of ϕ_R and ϕ_I can be defined analytically, but the Stokes set can generally only be determined by detailed examination of the topography of the integrand, which we shall now do.

5. The Stokes and anti-Stokes sets

We find the general topography of the integrand of (3) at particular points (x, y) outside the caustic by perturbing simple special cases. Since $C(x, y)$ is symmetric in y, we consider only $y \geqslant 0$. On the caustic at $x = -3(y/2)^{2/3}$ the double real stationary point of ϕ occurs at $\xi = +(-x/3)^{1/2}$, and the single real stationary point at $\xi = -2(-x/3)^{1/2}$. Using the conditions on the level lines of ϕ_I given previously, the level lines through the saddles of $|\exp(i\phi)|$, which are all at height one, can only be as sketched (as heavy full curves) in figure 2(a). Then the integration contour must be as shown by the broken and arrowed curve.

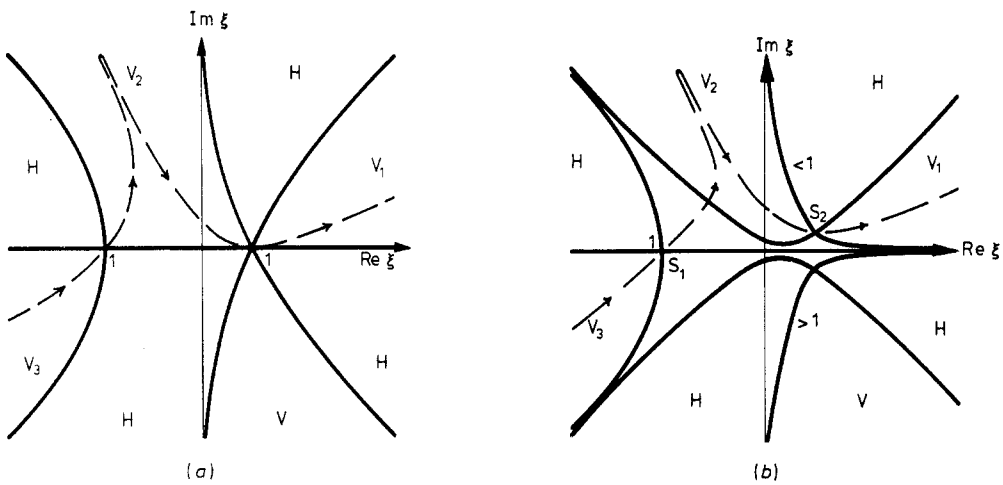

Figure 2. The complex ξ plane showing the level lines (heavy and full) of $|\exp(i\phi(\xi; x, y))|$ through the saddles, the integration contour (broken and arrowed), and the asymptotic hills H and valleys V. Heights of the level lines are indicated by <1, 1 and >1. The real ξ axis is always a level line: (a) is on the caustic, (b) is just outside the caustic, both for $y > 0$.

If we now move (x, y) outside the caustic a small distance, the double saddle splits into a complex conjugate pair of single saddles, as in figure 2(b). This lifts the degeneracy of the level lines, leaving one set at height one and producing sets at just above and just below one, associated with the two new complex saddles. (The relative heights of sets of level lines can be determined from the relative positions of their branches in the asymptotic cliff faces between valleys V and hills H.) The integration contour is carried upwards with the saddle S_2, and cannot pass through the lower saddle at all, so that only the real saddle and the upper complex saddle contribute to the integral. The integral is actually evaluated in this situation, using an approximation to the stationary points, in appendix C of BNW.

We can generalise the correspondence between real stationary phase points and real physical rays by saying that a *contributing complex stationary phase point* corresponds to a *complex ray*. So just outside the caustic there is one real ray and one complex ray. Keller (1958) introduced complex rays physically as an extension of geometrical optics, which was reviewed by Kravtsov (1967). The approach via steepest descent integration of the diffraction integral justifies Keller's physical assumption that only certain complex geometrical rays contribute to the diffraction pattern.

On the positive x axis the stationary points occur at $\xi = 0, \pm ix^{1/2}$. Now $\phi(\xi; x, 0)$ has a centre of symmetry at $\xi = 0$ (see equation (2)), making ϕ_I antisymmetric in ξ_R as well as in ξ_I. Hence, in this case only, the imaginary ξ axis is also a level line of $|\exp(i\phi)|$, and all the saddles and level lines through them are at height one. Therefore the level lines of $|\exp(i\phi)|$ and the integration contour must be as in figure 3(a). For a small positive y, a simple perturbation calculation shows that the saddles are shifted parallel to the real axis as shown (exaggerated) in figure 3(b). Again the degeneracy of the level lines is lifted, as indicated in the figure. By continuity with respect to y, the integration contour is little changed, as shown.

The important *topological* difference between figures 2(b) and 3(b) is that in figure 2(b) the line of steepest descent from saddle S_1 runs to valley V_2, whereas in figure 3(b) it runs to valley V_1. This changes the form of the integration contour, so the two figures must lie on opposite sides of a Stokes set. The saddle connection between saddle S_1 and the lower complex saddle is an example of a saddle connection which does not affect the

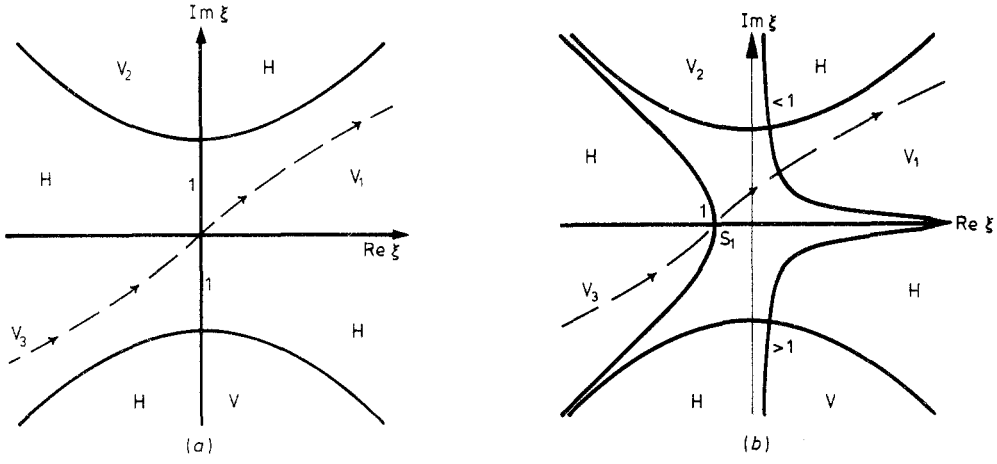

Figure 3. As figure 2 but (a) is on the positive x axis (here the imaginary ξ axis is also a level line), (b) is just above the positive x axis.

integration contour, although it occurs at the same time as the upper connection which does affect the contour.

Close to the positive x axis neither complex saddle contributes—there is one real ray only. This disappearance of the complex ray is a feature which is not apparent from Keller's geometrical approach, and one expects his method to give the wrong result in such regions. Somewhere between the caustic and the positive x axis the number of contributing complex saddles changes by one, which must occur along two lines running from the origin (the cusp point of the caustic) to infinity, symmetrically above and below the x axis. These lines form the Stokes set.

On crossing the Stokes set, the number of terms in the stationary phase approximation to the integral changes by one, and this is the reason for the name. In the theory of asymptotic expansions of functions of a complex variable, it is well known that the form of the asymptotic expansion changes discontinuously when the phase of the complex variable passes through certain values. This effect is called the Stokes phenomenon. It was discovered by Stokes (1864) while investigating the properties of Airy functions (essentially), which we now know are the simplest diffraction catastrophes. Dingle (1973) points out that this discontinuity of form is essential to preserve the numerical continuity of the analytic function being expanded, by cancelling the discontinuity in the numerical sum of one of the component infinite asymptotic series. The Stokes phenomenon occurs when the variable crosses certain lines radiating from the origin of the complex plane, called Stokes lines. If the function being asymptotically approximated has a contour integral representation, then the Stokes phenomenon corresponds to the number of stationary points contributing to the integral changing by one.

The Stokes set which we have introduced is a generalisation of the Stokes line to functions of many, possibly complex, variables defined by a contour integral such as (3). We may define the Stokes set generally to be the set of points in the (multidimensional complex) parameter space, other than the bifurcation set, at which the number of stationary points contributing to the integral changes by one.

Actually on the Stokes set, the topography of the integrand is intermediate between that shown in figures 2(b) and 3(b). The two branches of the integration contour which lie in valley V_2 of figure 2(b) coalesce when the line of steepest descent from saddle S_1 passes through saddle S_2, as indicated in figure 4. Approaching figure 4 from figure

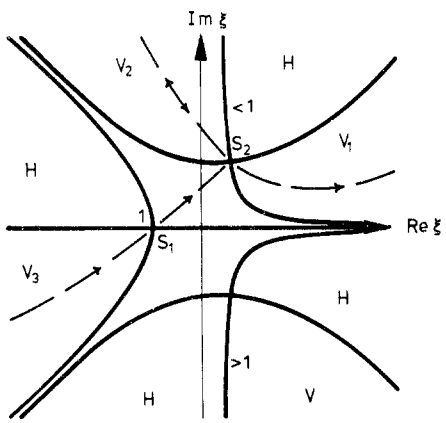

Figure 4. As figure 2, but on the Stokes set. The integration contour is indeterminate at saddle S_2.

2(*b*), the integration contour turns left through a right angle at S_2 and runs to infinity along V_2. It then returns along exactly the same path, as indicated by the arrows pointing in both directions, and passes straight over S_2 into V_1. The two branches of the contour in V_2 exactly cancel, so the contour is equivalent to that which turns to the right through a right angle at S_2 and passes directly into V_2. This is the contour which results as we approach figure 4 from figure 3(*b*).

This is the mechanism of the Stokes discontinuity, whereby the contribution of a saddle to the integral changes discontinuously. However, the integral itself depends continuously on its parameters, so any good approximation to it must be at least nearly continuous. Thus any discontinuity must be masked by a much larger continuous term, so that a saddle can only undergo a Stokes discontinuity when it is *subdominant* (e.g. Budden 1961, Ursell 1965) to (i.e. its amplitude is much less than) some other contribution to the integral (saddle S_1 in this example). This is part of the reason why the Stokes set is generally a subset of the saddle connection set. Our topographical analysis takes care of subdominance automatically.

The Maxwell set of ϕ_I determines the dominance of saddles, because on it saddles of $|\exp(i\phi)|$ occur at the same height. By analogy with anti-Stokes lines in conventional asymptotics (e.g. Budden 1961), we define the anti-Stokes set to be the set of points in the parameter space, other than the bifurcation set, on which saddles which may contribute to the integral exchange dominance. Therefore the anti-Stokes set is a subset of the Maxwell set of ϕ_I. It is not as important as the Stokes set, but it is useful, and appears naturally in any stationary phase analysis with complex stationary points. Since different saddles take part in a Stokes discontinuity on different branches of the Stokes set, there is usually a branch of the anti-Stokes set (or the bifurcation set) on which the saddles exchange dominance between branches of the Stokes set. This is certainly true for $C(x, y)$ with real (x, y), and for the Airy function of a complex argument. In fact, for the cusp diffraction catastrophe the anti-Stokes set interleaves between branches of the Stokes set in the same way that the conflict set interleaves between branches of the bifurcation set, on which different saddles coalesce.

In general, for real control parameters, the Stokes, anti-Stokes and bifurcation sets all have the same dimension, i.e. codimension one; e.g. for $C(x, y)$ they are all lines in the real (x, y) plane.

6. Equations of the Stokes and anti-Stokes sets

The cusp is probably the only diffraction catastrophe for which the equation of the Stokes set can be found explicitly, because of the lack of suitable formulae for the roots of a polynomial of degree greater than three. Suppose the real stationary point S_1 occurs at $\xi = \xi_r$, and the complex stationary point S_2 which may contribute occurs at $\xi = \xi_c$. On the Stokes set (figure 4) the line of steepest descent from ξ_r runs through ξ_c. The lines of steepest descent are level lines of ϕ_R, and those through ξ_r satisfy

$$\phi_R(\xi; x, y) = \phi_R(\xi_r; x, y).$$

On the Stokes set ξ_c lies on this curve, and the figures show that ξ_c can only lie on the correct branch as shown in figure 4. Remembering that the positions of the saddles depend on the parameters x and y, the equation of the Stokes set is given by

$$\phi_R[\xi_c(x, y); x, y] - \phi_R[\xi_r(x, y); x, y] = 0. \tag{6}$$

By solving explicitly for the stationary points, we show in appendix 1 that the equation of the Stokes set is

$$27y^2-(5+\sqrt{27})x^3=0.$$

(This can also be achieved without using explicit solutions of the cubic, but we need the explicit solutions in appendix 2). This is to be compared with the equation of the caustic:

$$27y^2+4x^3=0.$$

They are both semi-cubical parabolae, but the Stokes set is broader than the caustic and on the other side of the y axis, as shown in figure 5. It should be noted that this similarity between the Stokes and bifurcation sets is due to the simplicity of the problem—for higher diffraction catastrophes the Stokes set does not generally have the same form as the bifurcation set.

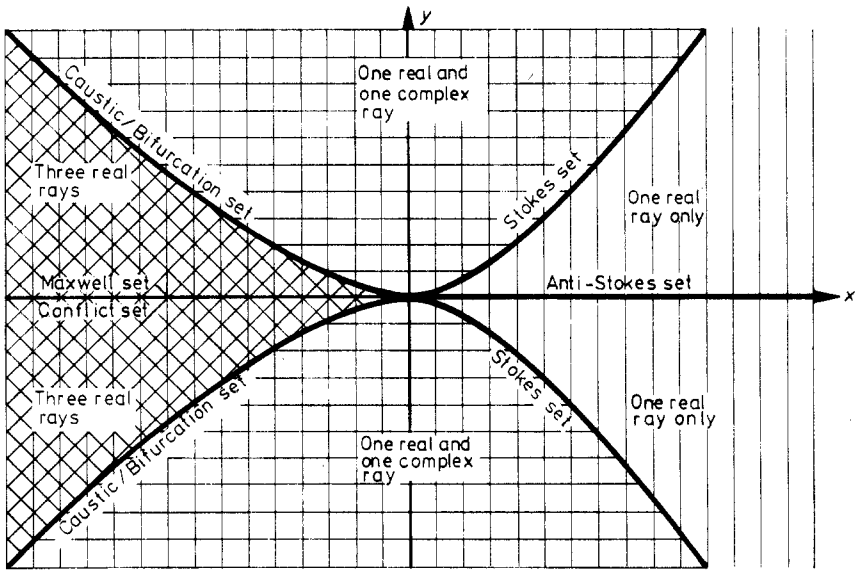

Figure 5. The regions of the control plane (x, y), having different numbers of rays or contributing stationary points, divided by the caustic and Stokes set. Also shown are the anti-Stokes and conflict sets.

The rest of the Maxwell set of ϕ_R is very easy to find, because equation (A1) of appendix 1 shows that the upper and lower saddles, which are a complex conjugate pair, *always* have the same value of ϕ_R. Hence the whole (x, y) plane outside the caustic is a Maxwell set for these two saddles. Between the caustic and the Stokes set (figure 2) they share a line of steepest descent, so all of this region is also a saddle-connection set of ϕ_R. To the right of the Stokes set (figure 3) the saddles of ϕ_R are not connected. The degeneracy of these sets (i.e. the fact that they are two-dimensional instead of one-dimensional) is due to the symmetry (5), namely

$$\phi(\xi^*; x, y)=\phi^*(\xi; x, y), \tag{5'}$$

of ϕ for real controls x, y.

The anti-Stokes set is much easier to find than the Stokes set. Moving in a circle about the origin in the (x, y) plane from the caustic to the positive x axis, the typical topography of the integrand is as shown successively in figures 2(a), 2(b), 4, 3(b), 3(a). Apart from on the bifurcation set (figure 2(a)), saddles only have the same amplitudes (i.e. one) on the positive x axis (figure 3(a)). This is the anti-Stokes set, as shown in figure 5. On crossing it into the lower half-plane $y \leq 0$, the lower saddle takes over the role of S_2 (because $\phi(\xi; x, -y) = \phi(-\xi; x, y)$, see equation (2)), i.e. S_2 exchanges dominance with S_1, which exchanges dominance simultaneously with the lower saddle, due again to the symmetry (5′) above.

7. Conclusion

We have introduced the Stokes and anti-Stokes sets of functions of many variables, defined by diffraction integrals. For the simplest diffraction catastrophe, the fold, alias the Airy integral function, the Stokes and anti-Stokes lines in the complex plane of its argument are well known. For the next member of the hierarchy, the cusp diffraction catastrophe $C(x, y)$, we have found the Stokes and anti-Stokes lines in the (x, y) plane of its two real arguments. The Stokes set is a broadened mirror image of the bifurcation set or caustic, a similarity which is largely fortuitous. The anti-Stokes set is the positive x axis, while the Maxwell set (which is also the conflict set) is the negative x axis. The caustic and the Stokes set partition the parameter space into regions containing different numbers of rays, as shown on figure 5 by the hatching. On crossing the caustic, two real rays turn into one complex ray; on crossing the Stokes set a complex ray appears or disappears.

Wavefront dislocations are phase singularities of complex diffraction wavefunctions. The Stokes set imposes a constraint upon their positions, because for dislocations to occur there must be at least two rays, so that they can destructively interfere. Therefore we expect no dislocations to the right of the Stokes set in figure 5, and this is borne out by exact computations. In fact, dislocations outside the caustic of the cusp diffraction catastrophe only occur in a single row very close to the caustic, because away from the caustic the amplitude of the complex ray decreases exponentially, so that it is no longer strong enough to cancel the real ray. In appendix 2 we make use of the exact positions of the stationary points found in appendix 1 to make a stationary phase approximation to $C(x, y)$ between the caustic and the y axis. From this we find the approximate positions of the wavefront dislocations.

Extension of the Stokes set of $C(x, y)$ into the complex parameter space should be worthwhile for the following reasons. The Airy function with complex arguments, used in a uniform approximation, is necessary to approximate a diffraction integral with two coalescing complex stationary points. This occurs, for example, in the swallowtail diffraction catastrophe with real controls, where the stationary points coalesce in two complex conjugate pairs (Wright 1977, Berry, Nye and Wright, in preparation). Similarly one expects $C(x, y)$ with complex arguments to be necessary to handle three coalescing complex stationary points, as will occur in the wigwam (A_6) diffraction catastrophe with real controls (where three conjugate pairs coalesce). Also complexification of control variables can transform diffraction integrals into partition functions of statistical mechanics. (See e.g. Güttinger and Eikemeier (1979, part V).)

Analysis of the Stokes set of the swallowtail diffraction catastrophe is in progress, although it is no longer possible to find explicit equations. Extension to umbilic

diffraction catastrophes, defined by double integrals, would be a major contribution to the unsolved problem of how to apply the method of steepest descent in four dimensions (Ursell 1980). Once the Stokes sets of the canonical diffraction catastrophe integrals are known, the Stokes set of a general diffraction catastrophe should be deducible via the method of uniform approximation (Berry 1976).

Acknowledgments

I thank the referees and my colleagues at Bristol for many helpful suggestions, and the Science Research Council for financial support.

Appendix 1. The equation of the Stokes set

Let $\xi = \xi_s$ at a stationary point, where s = r or c. Then ξ_s satisfies

$$\partial\phi(\xi_s; x, y)/\partial\xi = \xi_s^3 + x\xi_s + y = 0. \tag{A1}$$

We wish to evaluate

$$\phi(\xi_s; x, y) = \tfrac{1}{4}\xi_s^4 + \tfrac{1}{2}x\xi_s^2 + y\xi_s.$$

Subtracting $\frac{1}{4}\xi_s$ times equation (A1) (which is zero) gives

$$\phi(\xi_s; x, y) = \tfrac{1}{4}(x\xi_s^2 + 3y\xi_s),$$

so

$$\phi_R(\xi_s; x, y) = \tfrac{1}{4}[x(\xi_{sR}^2 - \xi_{sI}^2) + 3y\xi_{sR}], \tag{A2}$$

where the subscripts R and I indicate real and imaginary parts respectively. Using this result, equation (6) becomes

$$x\xi_r^2 + 3y\xi_r - x(\xi_{cR}^2 - \xi_{cI}^2) - 3y\xi_{cR} = 0. \tag{A3}$$

Now we must solve equation (A1) for ξ_r and ξ_c. We distinguish three cases: $x = 0$, $x > 0$ and $x < 0$.

$x = 0$ gives trivially

$$\xi_r = y^{1/3}, \qquad \xi_c = y^{1/3}\exp(\tfrac{1}{3}i\pi),$$

and the only solution of (A3) is $y = 0$.

For $x > 0$ we extend the standard trigonometric solution of a cubic (see e.g. Ferrar 1958) and write the required solutions of (A1) as

$$\xi_r = -2(\tfrac{1}{3}x)^{1/2}\sinh\theta, \tag{A4a}$$

$$\xi_c = (\tfrac{1}{3}x)^{1/2}(\sinh\theta + i\sqrt{3}\cosh\theta), \tag{A4b}$$

where

$$\sinh(3\theta) = \tfrac{1}{2}A, \qquad A = y(3/x)^{3/2}. \tag{A4c}$$

Substituting into (A3) gives

$$\sinh^2\theta + \cosh^2\theta - A\sinh\theta = 0, \tag{A5}$$

and eliminating $\sinh\theta$ between (A4c) and (A5) gives

$$A^4 - 10A^2 - 2 = 0,$$

whose real roots are

$$A = \pm(5+\sqrt{27})^{1/2}.$$

Therefore, since $A = y(3/x)^{3/2}$, the equation of the Stokes set for $x \geqslant 0$ is

$$27y^2 - (5+\sqrt{27})x^3 = 0.$$

For $x < 0$ the required solutions of (A1) may be written as

$$\xi_r = -2(-\tfrac{1}{3}x)^{1/2}\cosh\theta, \tag{A6a}$$

$$\xi_c = (-\tfrac{1}{3}x)^{1/2}(\cosh\theta + i\sqrt{3}\sinh\theta), \tag{A6b}$$

where

$$\cosh(3\theta) = \tfrac{1}{2}A, \qquad A = y(-3/x)^{3/2}. \tag{A6c}$$

The solution is $A = \pm(-5+\sqrt{27})^{1/2}$; but $(\sqrt{27}-5) < 4$, so that the curve given by this value of A would lie inside the caustic where the stationary points are all real. Therefore this curve is not part of the Stokes set.

Appendix 2. Wavefront dislocations outside the cusp caustic

In appendix 1 we found the stationary points outside the caustic exactly. We now use them to compute the positions of the dislocations in this region for $x < 0$, without making the perturbative approximation used by BNW. It is convenient to use A, instead of y, as a variable, and equations (A6) give the positions of the two contributing saddles. Then standard theory (e.g. Dingle 1973) gives the sum of the contributions from the real and complex saddles respectively as

$$C \simeq \frac{\exp\{i\tfrac{1}{4}[\pi - \tfrac{1}{3}x^2(4\cosh^2\theta + 2A\cosh\theta)]\}}{[-x(4\cosh^2\theta - 1)]^{1/2}}$$
$$+ \exp\{i\tfrac{1}{4}[-\pi + 2\tan^{-1}(\sqrt{3}\coth\theta) - \tfrac{1}{3}x^2(\cosh^2\theta - 3\sinh^2\theta - A\cosh\theta)]\}$$
$$\times \frac{\exp[\tfrac{1}{4}(x^2/\sqrt{3})(-2\cosh\theta + A)\sinh\theta]}{[2(-x)(4\cosh^2\theta - 1)^{1/2}\sinh\theta]^{1/2}}.$$

$C = 0$ gives the following two conditions, from considering respectively the amplitudes and phases of the two terms:

$$\frac{4\sinh^2\theta}{4\cosh^2\theta - 1} = \exp\left(\frac{-x^2}{\sqrt{3}}(A - 2\cosh\theta)\sinh\theta\right),$$

$$\tan^{-1}(\sqrt{3}\coth\theta) + (4n+1)\pi + \tfrac{1}{2}x^2(\cosh^2\theta + \sinh^2\theta + A\cosh\theta) = 0.$$

These two equations plus

$$A = y(-3/x)^{3/2} = 2\cosh(3\theta)$$

were solved numerically to give the results in the middle column of table A1. They are compared with the results quoted in appendix C of BNW. The right-hand column

Table A1. Positions of dislocations outside the cusp.

Perturbative stationary phase (equation C9) of BNW		Stationary phase (this paper)		Measured from phase plot (figure 14 of BNW) (estimated error ±0·05)	
x	y	x	y	x	y
−1·58	±1·70	−1·62	±1·68	−1·74	±1·65
−2·98	±3·19	−2·98	±3·16	−3·07	±3·13
−3·97	±4·40	−3·97	±4·39	−4·05	±4·35
−4·79	±5·50	−4·80	±5·49	−4·86	±5·45

shows values measured from the exact computed phase map of $C(x, y)$, and the left-hand column was calculated by a perturbative stationary phase method.

In every case the present results are closer than the perturbative results to the measured positions, but there is still a significant discrepancy. This is mainly because these dislocations are very close to the caustic, where the stationary points are not well separated and so the basic stationary phase approximation is becoming unreliable. This conclusion is supported by the fact that the discrepancy gets larger towards the cusp point, where all three stationary points coalesce.

The complexity of the present method, which gives only a slight increase in accuracy, emphasises the merit of the perturbative method. This applies particularly inside the caustic, where three stationary points contribute to the integral, and for the higher diffraction catastrophes which may have many more stationary points.

References

Abramowitz M and Stegun I A (eds) 1964 *Handbook of Mathematical Functions* (New York: Dover)

Airy G B 1838 *Trans. Camb. Phil. Soc.* **6** 379–402

Arnol'd V I 1973 *Usp. Mat. Nauk.* **28** (5) 17 (Engl. transl. 1973 *Russ. Math. Surv.* **28** (5) 19)

—— 1974 *Proc. Int. Congr. Math. Vancouver* pp 19–39 (quoted in Poston and Stewart 1978)

Berry M V 1976 *Adv. Phys.* **25** 1–26

Berry M V and Mackley M R 1977 *Phil. Trans. R. Soc.* A **287** 1–16

Berry M V, Nye J F and Wright F J 1979 *Phil. Trans. R. Soc.* A **291** 453–84

Berry M V and Upstill C 1980 *Catastrophe Optics* in *Progress in Optics* Vol XVIII, ed. E Wolf (Amsterdam: North-Holland) pp 257–346

Born M and Wolf E 1975 *Principles of Optics* (New York: Pergamon)

Budden K G 1961 *Radio Waves in the Ionosphere* (London: Cambridge UP)

Chester C, Friedman B and Ursell F 1957 *Proc. Camb. Phil. Soc.* **53** 599–611

Connor J N L 1976 *Mol. Phys.* **31** 33–55

Dennery P and Krzywicki A 1967 *Mathematics for Physicists* (New York: Harper International)

Dingle R B 1973 *Asymptotic Expansions: Their Derivation and Interpretation* (London: Academic)

Ferrar W L 1958 *Higher Algebra* (London: Oxford UP)

Güttinger W and Eikemeier H (eds) 1979 *Structural Stability in Physics* (Berlin: Springer)

Keller J B 1958 *Proc. Symp. Appl. Math.* **8** 27–52

Kravtsov Yu A 1967 *Izv. VUZ Radiofiz.* **10** 1283–304 (Engl. transl. Radiophys. Quantum Electron. **10** 719–30)

Nye J F and Berry M V 1974 *Proc. R. Soc.* A **336** 165–90

Pearcey T 1946 *Phil. Mag.* **37** 311–7

Poston T and Stewart I 1978 *Catastrophe Theory and Its Applications* (London: Pitman)

Stokes G G 1864 *Trans. Camb. Phil. Soc.* **10** 106–28

Trinkaus H and Drepper F 1977 *J. Phys. A: Math. Gen.* **10** L11–6
Thom R 1975 *Structural Stability and Morphogenesis* (Reading, MA: Benjamin)
Ursell F 1965 *Proc. Camb. Phil. Soc.* **61** 113–28
—— 1980 *Math. Proc. Camb. Phil. Soc.* **87** 249–73
Wright F J 1977 *PhD thesis* Bristol University
Wright F J 1979 in *Structural Stability in Physics* ed. W Güttinger and H. Eickemeier (Berlin: Springer) pp 141–56