Chapter One

Introduction

> Probably no mathematical structure is richer, in terms of the variety of physical situations to which it can be applied, than the equations and techniques that constitute wave theory. Eigenvalues and eigenfunctions, Hilbert spaces and abstract quantum mechanics, numerical Fourier analysis, the wave equations of Helmholtz (optics, sound, radio), Schrödinger (electrons in matter) ... variational methods, scattering theory, asymptotic evaluation of integrals (ship waves, tidal waves, radio waves around the earth, diffraction of light)—examples such as these jostle together to prove the proposition.
>
> *M. V. Berry* [1]

> There is a theory which states that if ever anyone discovers exactly what the Universe is for and why it is here, it will instantly disappear and be replaced by something even more bizarre and inexplicable. There is another theory which states that this has already happened.
>
> *Douglas Adams* [155]

Douglas Adams's famous *Hitchhiker* trilogy consists of *five* books; coincidentally this book addresses the three topics of rays, waves, and scattering in five parts: (i) Rays, (ii) Waves, (iii) Classical Scattering, (iv) Semiclassical Scattering, and (v) Special Topics in Scattering Theory (followed by six appendices, some of which deal with more specialized topics). I have tried to present a coherent account of each of these topics by separating them insofar as it is possible, but in a very real sense they are inseparable. We are in effect viewing each phenomenon (e.g. a rainbow) from several different structural directions or different mathematical levels of description. It is tempting to regard descriptions and explanations as synonymous, but of course they are not. In fact, regarding the rainbow in particular

> Aristotle and later scientists in antiquity "constructed theories that primarily *describe* natural phenomena in mathematical or geometric terms, with little or no concern for physical mechanisms that might *explain* them." This contrast goes to the heart of the difference between "Aristotelian" and mathematical modeling. [169]

The first two chapters of Part I (Rays) introduce the topic from a physical and a mathematical point of view, respectively. Thereafter the subject of rays is viewed in an atmosphere–sea–earth sequence, merging with Part II (Waves) via the reverse sequence earth–sea–atmosphere. Part III (Classical Scattering) examines the relationship between the elegance of the classical Lagrangian and Hamiltonian formulations of mechanics and optics, and having set the scene, so to speak, moves on to develop Kepler's laws of planetary motion (arising from what I refer to as gravitational scattering). This is followed by a revisitation of the topics of surface gravity waves, acoustics, electromagnetic scattering (including the Mie solution), and diffraction. Part IV (Semiclassical Scattering) provides a transition from Part III and addresses more of the nuts and bolts of the underlying mathematics. In so doing it allows us to reexamine the WKB(J) approximation and apply it to some simple one-dimensional potentials. Readers may wonder at my notation for this approximation—why WKB(J)? It is to pay homage to the legacy of Sir Harold Jeffreys in this and many other areas of applied mathematics and mathematical physics. As an eighteen-year-old student I bought a copy of the celebrated *Methods of Mathematical Physics* [98], co-written with his wife, Lady Bertha Jeffreys. I vowed that I would try to understand as much of it as possible (and I have tried). It is a magnificent book. But writing WKBJ might be a bridge too far, as they say—the majority of citations in the literature appear to use the WKB form, so I have adapted the acronym accordingly. And it appears that I am not the only one to think Jeffreys deserves more credit than he has been given: see the first quotation in Chapter 22.

Writing the final chapter in this part was a fascinating undertaking; it is a collection of salient properties of Sturm-Liouville systems with particular reference to the time-independent Schrödinger equation. It touches on several aspects of the theory of differential equations that have profound implications for the topics in this book and is distilled from a variety of sources (some of which are now quite difficult to find). I hope that the reader will benefit from having these all in one place. Part V (Special Topics in Scattering Theory) is an eclectic anthology of material that has intrigued me over the years, and some of the topics are reflections of joint research with recent graduate students. Six appendices round out the material, the second of which is a condensed version of a chapter in [68] (see also [236]). Appendix D is also based on my contribution to a joint article written with former students [149].

I should also say something about the quotations I have chosen. At the beginning of most chapters I have placed a quotation (and occasionally more than one) from a book, scientific paper, or internet article that I consider to be a brief but pertinent introduction to the subject matter of that chapter—an appetite-whetter if you will. I have even cited material from Wikipedia (gasp!). In some of the chapters I have merely waffled on for a bit instead of using an introductory quotation. Regarding short excerpts from scientific articles on the internet, in the few cases for which I have been unable to find the name of the author (after extensive searching), I trust that the link provided will suffice for the interested reader to pursue the topic (and author) further.

In the Preface I mentioned that this book is an intensely personal one, and how the beautiful phenomenon of a rainbow serves as an implicit template or directory for much of the rest of the book. It does so because much of the motivation for the design of the book springs from it, as evidenced below. And as to *why* I have chosen to structure things in this way, you will need to read the first paragraph of the last chapter ("Back where we started") to find out!

1.1 THE RAINBOW DIRECTORY

> Optical phenomena visible to everyone have been central to the development of, and abundantly illustrate, important concepts in science and mathematics. The phenomena considered from this viewpoint are rainbows, sparkling reflections on water, mirages, green flashes, earthlight on the moon, glories, daylight, crystals and the squint moon. And the concepts involved include refraction, caustics (focal singularities of ray optics), wave interference, numerical experiments, mathematical asymptotics, dispersion, complex angular momentum (Regge poles), polarisation singularities, Hamilton's conical intersections of eigenvalues ('Dirac points'), geometric phases and visual illusions. [151]

> The theory of the rainbow has been formulated at many levels of sophistication. In the geometrical-optics theory of Descartes, a rainbow occurs when the angle of the light rays emerging from a water droplet after a number of internal reflections reaches an extremum. In Airy's wave-optics theory, the distortion of the wave front of the incident light produced by the internal reflections describes the production of the supernumerary bows and predicts a shift of a few tenths of a degree in the angular position of the rainbow from its geometrical-optics location. In Mie theory, the rainbow appears as a strong enhancement in the electric field scattered by the water droplet. Although the Mie electric field is the exact solution to the light-scattering problem, it takes the form of an infinite series of partial-wave contributions that is slowly convergent and whose terms have a mathematically complicated form. In the complex angular momentum theory, the sum over partial waves is replaced by an integral, and the rainbow appears as a confluence of saddle-point contributions in the portion of the integral that describes light rays that have undergone *m* internal reflections within the water droplet. [168]

1.1.1 The Multifaceted Rainbow

What follows is a partial list of context-useful descriptions of a rainbow; they are certainly not mutually exclusive categories. References are made to (italicized) topics that will be expanded in later chapters, so the reader should not be unduly concerned about words or phrases in this Chapter that have yet to be defined (e.g., as in the above quotations). Such topics will be unfolded in due course, so be patient, dear reader. But for those who wish to know *now* whether "the Butler did it," and if not, who did, the answers *are* in the back of the book, and it is recommended that you turn immediately to the short last chapter, titled "Back Where We Started" prior to returning here.

In part, a rainbow is:

(1) A concentration of light rays corresponding to an extremum of the scattering angle $D(i)$ as a function of the angle of incidence i. In particular it is a minimum for the primary bow, and the exiting ray at this minimum value is called the *Descartes or rainbow ray*; as noted below, the notation $D(i)$ is replaced by $\theta\,(b)$ in most of the scattering literature, where b is the *impact parameter*. The angle θ is much used in connection with the equations of scattering in the chapters that follow, but

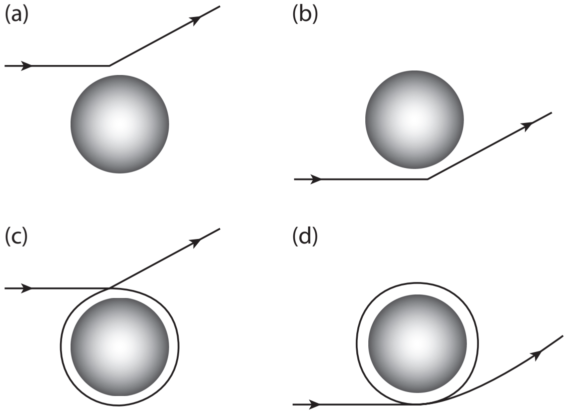

Figure 1.1 (a) $\Theta = \theta$; (b) $\Theta = -\theta$; (c) $\Theta = \theta - 2\pi$; (d) $\Theta = -\theta - 2\pi$.

the former notation will be retained when discussing the optical rainbow (I trust this will be at worst only a minor annoyance for my colleagues in meteorological optics);

(2) An integral superposition of waves over a (locally) cubic wavefront (*the Airy approximation*). This approximation is valid only for large size parameters (i.e., the drop size is much larger than the wavelength of the incident light) and for small deviations from the angle of the rainbow ray; and relatedly

(3) An interference phenomenon within the same light wave (the origin of the *supernumerary bows*);

(4) A *caustic*, separating a 2-ray region from a 0-ray (or shadow) region; and relatedly

(5) A *fold diffraction catastrophe*.

(6) Related to the behavior of the *scattering amplitude* for the third term in what is referred to as a *Debye series expansion*. The dominant contribution to the third Debye term in the rainbow region is a *uniform asymptotic expansion*. On the bright side of the rainbow, this result matches smoothly with the *WKB(J) approximation* in the 2-ray region. On the dark side, it does so again, this time with the damped complex *saddle-point contribution*. Put differently, diffraction into the shadow side of a rainbow occurs by *tunneling*. And relatedly

(7) A coalescence of two *real saddle points*;

(8) A result of ray/wave interactions/tunneling with an *effective potential* comprising a *square well* and a *centrifugal barrier*.

Deflection Angle

The *classical deflection angle* is denoted by $\Theta(b)$, where b is the impact parameter. The scattering angle is $\theta(b)$, and the two angles are related by

$$\Theta(b) + 2n\pi = \pm\theta, \tag{1.1}$$

where n is a non-negative integer chosen such that $\theta \in [0, \pi]$. Illustrative examples of the relationship (1.1) for the *same value* of θ are shown in Figure 1.1. For a repulsive potential (as in Figure 1.1a), $\Theta = \theta$, but in principle for an attractive potential Θ can be arbitrarily negative, because the particle (or ray) may orbit the scattering center many times before emerging from it. Thus there may be several different trajectories that lead to the same scattering angle θ (this will be discussed further in Chapter 16).

1.2 A MATHEMATICAL TASTE OF THINGS TO COME

1.2.1 Rays

The Ray-Theoretic Rainbow

A rainbow occurs when the the scattering angle $D(i)$, as a function of the angle of incidence i, passes through an extremum. The folding back of the corresponding scattered or deviated ray takes place at this extremal scattering angle (the rainbow angle $D_{\min} = \theta_R$; note that sometimes in the popular literature the rainbow angle is loosely interpreted as the *supplement* of the deviation, $\pi - D_{\min}$). Two rays scattered in the same direction with different angles of incidence on the illuminated side of the rainbow ($D > \theta_R$) fuse together at the rainbow angle and disappear as the "dark side" ($D < \theta_R$) is approached. This is one of the simplest physical examples of a *fold catastrophe* in the sense of Thom, as will be discussed later. It will also be shown later that rainbows of different orders can be associated with so-called *Debye terms* of different orders; the primary and secondary bows correspond to $p = 2$ and $p = 3$, respectively, where $p > 1$ is the number of ray paths inside the drop, so the number of *internal reflections* is $k = p - 1$.

Details

For $p - 1$ such internal reflections the total deviation is

$$D_{p-1}(i) = (p-1)\pi + 2(i - pr), \tag{1.2}$$

where by Snell's law, $r = r(i)$. This approach can be thought of as the elementary classical description.

For the primary rainbow ($p = 2$),

$$D_1(i) = \pi - 4r(i) + 2i, \tag{1.3}$$

and for the secondary rainbow ($p = 3$),

$$D_2 = 2i - 6r(i) \tag{1.4}$$

(modulo 2π). For $p = 2$ the angle through which a ray is deviated is

$$D_1(i) = \pi + 2i - 4\arcsin\left(\frac{\sin i}{n}\right), \tag{1.5}$$

where n is now the relative refractive index of the medium. In general

$$D_{p-1}(i) = (p-1)\pi + 2i - 2p \arcsin\left(\frac{\sin i}{n}\right), \tag{1.6}$$

or equivalently,

$$D_k(i) = k\pi + 2i - 2(k+1) \arcsin\left(\frac{\sin i}{n}\right). \tag{1.7}$$

For $p = 2$ the minimum angle of deviation occurs when

$$i = i_c = \arccos\left(\frac{n^2-1}{3}\right)^{1/2}, \tag{1.8}$$

and more generally when

$$i = i_c = \arccos\left(\frac{n^2-1}{p^2-1}\right)^{1/2}, \tag{1.9}$$

(this clearly places constraints on both n and p if bows are to occur). For the primary rainbow $i_c \approx 59°$, using an approximate value for water of $n = 4/3$; from this it follows that $D_1(i_c) = D_{\min} = \theta_R \approx 138°$. For the secondary bow $i_c \approx 72°$; and $D_2(i_c) = \theta_R \approx 129°$. For $p = 2$, in terms of n alone, $D_1(i_c) = \theta_R$ (the rainbow angle) is defined by

$$D_1(i_c) = \theta_R = 2 \arccos\left[\frac{1}{n^2}\left(\frac{4-n^2}{3}\right)^{3/2}\right]. \tag{1.10}$$

1.2.2 Waves

Surprising as it may seem to anyone who thinks in terms of rays alone, there is also a wave-theoretic approach to the rainbow problem. The essential mathematical problem for scalar waves can be thought of either in classical terms (e.g., the scattering of sound waves) or in wave-mechanical terms (e.g., the nonrelativistic scattering of particles by a square potential well (or barrier) of radius a and depth (or height) V_0). In either case we can consider a scalar plane wave in spherical geometry impinging in the direction $\theta = 0$ on a penetrable ("transparent") sphere of radius a. The wave function $\psi(r)$ satisfies the scalar Helmholtz equation

$$\nabla^2\psi + n^2k^2\psi = 0, \quad r < a; \tag{1.11}$$

$$\nabla^2\psi + k^2\psi = 0, \quad r \geq a, \tag{1.12}$$

where $n > 1$ is the refractive index of the sphere (a similar problem can be posed for gas bubbles in a liquid, for which $n < 1$). The boundary conditions are that $\psi(r)$ and $\psi'(r)$ are continuous at the surface.

1.2.3 Scattering (Classical)

In what follows θ is the angle of observation, measured from the forward to the scattered direction, and thus defines the *scattering plane*. At large distances from the sphere ($r \gg a$) the wave field ψ can be decomposed into an incident wave + scattered field:

$$\psi \sim e^{ikr\cos\theta} + \frac{f(k,\theta)}{r}e^{ikr}, \tag{1.13}$$

where the scattering amplitude $f(k,\theta)$ (which can be made dimensionless with respect to the radius a) is defined as

$$f(k,\theta) = \frac{1}{2ik}\sum_{l=0}^{\infty}(2l+1)(S_l(k)-1)P_l(\cos\theta). \tag{1.14}$$

This is the Faxen-Holtzmark formula, and it will be encountered in several different guises later in the book (e.g., Section 21.5). $S_l(k)$ is an element of the scattering matrix (or function or operator, depending on context)) for a given l, and P_l is a Legendre polynomial of degree l. But what *is* the S-matrix? In optical terms, it is the partial-wave scattering amplitude with diffraction omitted—of course, this just kicks the can down the road—what is the partial-wave scattering amplitude? What (really) is diffraction? The reader's forbearance is requested; for now float in a sea of relatively undefined terms and just soak. Fundamentally, the S-matrix is the portion of the partial-wave scattering amplitudes that corresponds to a direct interaction between the incoming wave and the scattering particle. In very simplistic terms it converts an ingoing wave to an outgoing one, and for a spherical square well or barrier we shall find that

$$S_l = -\frac{h_l^{(2)}(\beta)}{h_l^{(1)}(\beta)}\left\{\frac{\ln' h_l^{(2)}(\beta) - n\ln' j_l(\alpha)}{\ln' h_l^{(1)}(\beta) - n\ln' j_l(\alpha)}\right\}, \tag{1.15}$$

where, following the notation of [26], $\ln'$ represents the logarithmic derivative operator, and j_l and h_l are spherical Bessel and Hankel functions, respectively. $\beta = 2\pi a/\lambda \equiv ka$ is the dimensionless external wavenumber, and $\alpha = n\beta$ is the corresponding internal wavenumber. S_l may be equivalently expressed in terms of cylindrical Bessel and Hankel functions. The lth partial wave in the *Mie solution* is associated with an impact parameter

$$b_l = \left(l + \frac{1}{2}\right)k^{-1}, \tag{1.16}$$

that is, only rays hitting the sphere ($b_l \lesssim a$) are significantly scattered, and the number of terms that must be retained in the Mie series to get an accurate result is of order β. As implied earlier, for visible light scattered by water droplets in the atmosphere, $\beta \sim$ several thousand. This is why, to quote Arnold Sommerfeld [170]:

> The electromagnetic study of light diffraction on an object is a very complicated problem even in the case of the sphere, the simplest possible one. The field outside a sphere can be represented by series of spherical harmonics and Bessel functions of half-integer indices. These series have been discussed by G. Mie for colloidal particles of arbitrary compositions. But even there a mathematical difficulty develops which quite generally is a drawback of this method of series development: for fairly large particles ($\beta = ka$, $a =$ radius, $k = 2\pi/\lambda$) the series converge so slowly that they become practically useless. Except for this difficulty, we could, in this way, obtain a complete solution of the problem of the rainbow.

This problem can be remedied by using the *Poisson summation formula* (related to the *Watson transform*) to rewrite $f(k, \theta)$ in terms of the integral

$$f(\beta, \theta) = \frac{i}{\beta} \sum_{m=-\infty}^{\infty} (-1)^m \int_0^{\infty} [1 - S(\lambda, \beta)] P_{\lambda - \frac{1}{2}}(\cos\theta) e^{2im\pi\lambda} \lambda d\lambda. \tag{1.17}$$

For fixed β, $S(\lambda, \beta)$ is a meromorphic function of the complex variable $\lambda = l + 1/2$, which should not be confused with the wavelength (the context should always make this distinction clear). In particular in what follows it is the *poles* of this function that are of interest. In terms of cylindrical Bessel and Hankel functions, they are defined by the condition

$$\ln' H_\lambda^{(1)}(\beta) = n \ln' J_\lambda(\alpha), \tag{1.18}$$

and are called *Regge poles* in the scattering theory literature (see, e.g., [51]). Typically they are associated with surface waves for the *impenetrable* sphere problem; for the *transparent* sphere two types of Regge poles arise—one type leading to rapidly convergent residue series (diffracted or creeping rays), and the other type associated with resonances via the internal structure of the potential. Many of these are clustered close to the real axis, spoiling the rapid convergence of the residue series. Mathematically, the resonances are complex eigenfrequencies associated with the poles λ_j of the scattering function $S(\lambda, k)$ in the first quadrant of the complex λ-plane; they are known as Regge poles (for real β). The imaginary parts of the poles are directly related to resonance widths (and therefore lifetimes). As the index j decreases, $\operatorname{Re}\lambda_j$ increases and $\operatorname{Im}\lambda_j$ decreases very rapidly (reflecting the exponential behavior of the barrier transmissivity). As β increases, the poles λ_j trace out Regge trajectories, and $\operatorname{Im}\lambda_j$ tends exponentially to zero. When $\operatorname{Re}\lambda_j$ passes close to a “physical” value, $\lambda = l + 1/2$, it is associated with a resonance in the lth partial wave; the larger the value of β, the sharper the resonance becomes for a given node number j.

In [83] it is shown that

$$S(\lambda, \beta) = \frac{H_\lambda^{(2)}(\beta)}{H_\lambda^{(1)}(\beta)} \left(R_{22}(\lambda, \beta) + T_{21}(\lambda, \beta) T_{12}(\lambda, \beta) \frac{H_\lambda^{(1)}(\alpha)}{H_\lambda^{(2)}(\alpha)} \sum_{p=1}^{\infty} [\rho(\lambda, \beta)]^{p-1} \right), \tag{1.19}$$

where

$$\rho(\lambda, \beta) = R_{11}(\lambda, \beta) \frac{H_\lambda^{(1)}(\alpha)}{H_\lambda^{(2)}(\alpha)}. \tag{1.20}$$

This is the *Debye expansion*, arrived at by expanding the expression $[1 - \rho(\lambda, \beta)]^{-1}$ as an infinite geometric series. R_{22}, R_{11}, T_{21}, and T_{12} are, respectively, the external/internal reflection and internal/external transmission coefficients for the problem. This procedure transforms the interaction of wave + sphere into a series of surface interactions. In so doing it unfolds the stationary points of the integrand, so that a given integral in the Poisson summation contains at most one stationary point. This permits a ready identification of the many terms in accordance with ray theory. The first term inside the parentheses represents direct reflection from the surface. The pth term represents transmission into the sphere (via the term T_{21}) subsequently bouncing back and forth between $r = a$ and $r = 0$ a total of p times with $p - 1$ internal reflections at the surface (this time via the R_{11}term in ρ). The middle factor in the second term, T_{12}, corresponds to transmission to the outside medium. In general, therefore, the pth term of the Debye expansion represents the effect of $p + 1$ surface interactions.

1.2.4 Scattering (Semiclassical)

On the way to the scattering representation, so to speak, there is a *semiclassical* description. In a primitive sense, the semiclassical approach is the ‘geometric mean’ between classical and quantum mechanical descriptions of phenomena in which interference and diffraction effects enter the picture. The latter do so via the transition from geometrical optics to wave optics. This is a characteristic feature of the ‘primitive’ semiclassical formulation. Indeed, the infinite intensities predicted by geometrical optics at focal points, lines and caustics in general are “breeding grounds” for diffraction effects, as are light/shadow boundaries for which geometrical optics predicts finite discontinuities in intensity. Such effects are most significant when the wavelength is comparable with (or larger than) the typical length scale for variation of the physical property of interest (e.g. size of the scattering object). Thus a scattering object with a “sharp” boundary (relative to one wavelength) can give rise to *diffractive scattering* phenomena.

There are ‘critical’ angular regions where the primitive semiclassical approximation breaks down, and diffraction effects cannot be ignored, although the angular ranges in which such critical effects become significant get narrower as the wavelength decreases. Early work in this field contained *transitional asymptotic approximations* to the scattering amplitude in these ‘critical’ angular domains, but they have very narrow domains of validity, and do not match smoothly with neighboring

> 'non-critical' angular domains. It is therefore of considerable importance to seek *uniform asymptotic approximations* that by definition do not suffer from these failings. Fortunately, the problem of plane wave scattering by a homogeneous sphere exhibits all of the critical scattering effects (and it can be solved exactly, in principle), and is therefore an ideal laboratory in which to test both the efficacy and accuracy of the various approximations. Furthermore, it has relevance to both quantum mechanics (as a square well or barrier problem) and optics (Mie scattering); indeed, it also serves as a model for the scattering of acoustic and elastic waves, and was studied in the early twentieth century as a model for the diffraction of radio waves around the surface of the earth. [88]

It transpires that the integral for $f(\beta, \theta)$ can be expressed as an infinite sum:

$$f(\beta, \theta) = f_0(\beta, \theta) + \sum_{p=1}^{\infty} f_p(\beta, \theta), \tag{1.21}$$

where

$$f_0(\beta, \theta) = \frac{i}{\beta} \sum_{m=-\infty}^{\infty} (-1)^m \int_0^{\infty} \left[1 - \frac{H_\lambda^{(2)}(\beta)}{H_\lambda^{(1)}(\beta)} R_{22} \right] P_{\lambda-\frac{1}{2}}(\cos\theta) \exp(2im\pi\lambda)\lambda d\lambda. \tag{1.22}$$

The expression for $f_p(\beta, \theta)$ involves a similar type of integral for $p \geq 1$. The application of the modified Watson transform to the third term ($p = 2$) in the Debye expansion of the scattering amplitude shows that it is this term which is associated with the phenomena of the primary rainbow. More generally, for a Debye term of given order p, a rainbow is characterized in the λ-plane by the occurrence of two real saddle points $\bar{\lambda}$ and $\bar{\lambda}'$ between 0 and β in some domain of scattering angles θ, corresponding to the two scattered rays on the light side. As $\theta \to \theta_R^+$ the two saddle points move toward each other along the real axis, merging together at $\theta = \theta_R$. As θ moves into the dark side ($\theta < \theta_R$), the two saddle points become complex, moving away from the real axis in complex conjugate directions. Therefore, from a mathematical point of view, *a rainbow can be defined as a coalescence of two saddle points in the complex angular momentum plane* (see Figure 1.2).

Thus the rainbow light/shadow transition region is associated physically with the confluence of a pair of geometrical rays and their transformation into complex rays; mathematically this corresponds to a pair of real saddle points merging into a complex saddle point. Then the problem is to find the asymptotic expansion of an integral having two saddle points that move toward or away from each other. The generalization of the standard saddle-point technique to include such problems was made by Chester et al. [164] and using their method, Nussenzveig [84] was able to find a uniform asymptotic expansion of the scattering amplitude that was valid throughout the rainbow region and that matched smoothly onto results for neighboring regions. The lowest order approximation in this expansion turns out to be the celebrated *Airy approximation*, which, despite several attempts to improve on it, was the best approximate treatment prior to the analyses of Nussenzveig and coworkers. However, Airy's theory had a limited range of applicability as a result

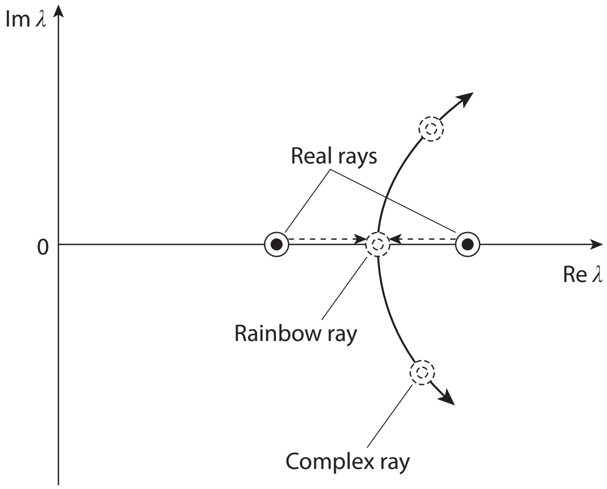

Figure 1.2 The coalescence of two real saddle points in the complex λ-plane as the rainbow angle is approached.

of its underlying assumptions; by contrast, the uniform expansion is valid over a much larger range.

1.2.5 Caustics and Diffraction Catastrophes

An alternative way of describing the rainbow phenomenon is by way of catastrophe theory, the rainbow being one of the simplest examples in catastrophe optics. Before summarizing the mathematical details of the basic rainbow diffraction catastrophe, it will be useful to introduce some of the language used in catastrophe-theoretic arguments. As we know from geometrical optics the primary bow deviation angle D_1 (in particular) has a minimum corresponding to the rainbow angle (or Descartes ray) $D_{\min} = \theta_R$ when considered as a function of the angle of incidence i. Clearly the point $(i, D_1(i))$ corresponding to this minimum is a singular point (approximately $(59^\circ, 138^\circ)$) insofar as it separates a two-ray region ($D_1 > D_{\min}$) from a zero-ray region ($D_1 < D_{\min}$) in the geometrical-optics level of description. This is a singularity or *caustic point.* The rays form a directional caustic at this point, and this is a *fold catastrophe* (symbol: A_2), the simplest example of a catastrophe. It is the only stable singularity with *codimension* one (the dimensionality of the control space (one) minus the dimensionality of the singularity itself, which is zero). In space the caustic surface is asymptotic to a cone with semi-angle 42°.

Optics is concerned to a great extent with families of rays filling regions of space; the *singularities* of such ray families are *caustics.* For optical purposes this level of description is important for classifying caustics using the concept of *structural stability*: this enables one to classify those caustics whose topology survives perturbation. Structural stability means that if a singularity S_1 is produced by a generating function ϕ_1 (see below for an explanation of these terms), and ϕ_1 is perturbed to

ϕ_2, the correspondingly changed S_2 is related to S_1 by a *diffeomorphism* of the control set C (that is by a smooth reversible set of control parameters; in other words a smooth deformation). In the present context this means, in physical terms, that distortions of incoming wavefronts by deviations of the raindrop shapes from their ideal spherical form does *not* prevent the formation of rainbows, though there may be some changes in the features. Another way of expressing this concept is to describe the system as well-posed in the limited sense that small changes in the input generate correspondingly small changes in the output. For the so-called elementary catastrophes, structural stability is a generic (or typical) property of caustics. Each structurally stable caustic has a characteristic diffraction pattern, the wave function of which has an integral representation in terms of the standard polynomial describing the corresponding catastrophe. From a mathematical point of view these diffraction catastrophes are especially interesting, because they constitute a new hierarchy of functions, distinct from the special functions of analysis. A review of this subject has been made by Berry and Upstill [102] wherein may be found an introduction to the formalism and methods of catastrophe theory as developed particularly by Thom [171] but also by Arnold [173]. The books by Gilmore [105] (Chapter 13 of which concerns caustics and diffraction catastrophes) and Poston and Stewart [172] are noteworthy in that they also provide many applications.

Diffraction can be discussed in terms of the scalar Helmholtz equation in some spatial region $\mathcal{R}$,

$$\nabla^2 \psi(\mathcal{R}) + k^2 n^2(\mathcal{R}) \psi(\mathcal{R}) = 0, \tag{1.23}$$

for the complex scalar wavefunction $\psi(\mathcal{R})$, k being the free-space wavenumber and n the refractive index. The concern in catastrophe optics is to study the asymptotic behavior of wave fields near caustics in the short-wave limit $k \to \infty$ (semiclassical theory). In a standard manner, $\psi(\mathcal{R})$ is expressed as

$$\psi(\mathcal{R}) = a(\mathcal{R}) e^{ik\chi(\mathcal{R})}, \tag{1.24}$$

where the modulus a and the phase $k\chi$ are both real quantities. To the lowest order of approximation χ satisfies the *Hamilton-Jacobi equation*, and ψ can be determined asymptotically in terms of a phase-action exponent (surfaces of constant action are the wavefronts of geometrical optics). The integral representation for ψ is

$$\psi(\mathcal{R}) = e^{-in\pi/4} \left(\frac{k}{2\pi} \right)^{n/2} \int \dots \int b(s; \mathcal{R}) \exp[ik\phi(s; \mathcal{R})] d^n s, \tag{1.25}$$

where n is the number of state (or behavior) variables s, and b is a weight function. In general there is a relationship between this representation and the simple ray approximation [102]. According to the principle of stationary phase, the main contributions to the above integral for given $\mathcal{R}$ come from the stationary points (i.e., those points s_i for which the gradient map $\partial\phi/\partial s_i$ vanishes) caustics are *singularities* of this map, where two or more stationary points coalesce. Because $k \to \infty$, the integrand is a rapidly oscillating function of s, so other than near the points s_i, destructive interference occurs and the corresponding contributions are negligible. The stationary points are well separated, provided $\mathcal{R}$ is not near a caustic; the simplest form of stationary phase can then be applied and yields a series of terms of the form

$$\psi(\mathcal{R}) \approx \sum_{\mu} a_{\mu} \exp[i g_{\mu}(k, \mathcal{R})], \tag{1.26}$$

where the details of the g_μ need not concern us here. Near a caustic, however, two or more of the stationary points are close (in some appropriate sense), and their contributions cannot be separated without a reformulation of the stationary phase principle to accommodate this [164]. The problem is that the "ray" contributions can no longer be considered separately; when the stationary points approach closer than a distance $O(k^{-1/2})$, the contributions are not separated by a region in which destructive interference occurs (See Appendix A for the "order" nomenclature used this book). When such points coalesce, $\phi(s; \mathcal{R})$ is stationary to higher than first order, and quadratic terms as well as linear terms in $s - s_\mu$ vanish. This implies the existence of a set of displacements ds_i, away from the extrema s_μ, for which the gradient map $\partial\phi/\partial s_i$ still vanishes, that is, for which

$$\sum_i \frac{\partial^2 \phi}{\partial s_i \partial s_j} ds_i = 0. \tag{1.27}$$

The condition for this homogeneous system of equations to have a solution (i.e., for the set of control parameters X to lie on a caustic) is that the Hessian

$$H(\phi) \equiv \det\left(\frac{\partial^2 \phi}{\partial s_i \partial s_j}\right) = 0, \tag{1.28}$$

at points $s_\mu(X)$ where $\partial\phi/\partial s_i = 0$ (details can be found in [102]). The caustic defined by $H = 0$ determines the bifurcation set for which at least two stationary points coalesce (in the present circumstance this is just the rainbow angle). In view of this discussion there are two other ways of expressing this: (i) rays coalesce on caustics, and (ii) caustics correspond to singularities of gradient maps. To remedy this problem the function ϕ is replaced by a simpler "normal form" Φ with the same stationary-point structure; the resulting diffraction integral is evaluated exactly. This is where the property of structural stability is so important, because if the caustic is structurally stable it must be equivalent to one of the catastrophes (in the diffeomorphic sense described above). The result is a generic diffraction integral that will occur in many different contexts. The basic diffraction catastrophe integrals (one for each catastrophe) may be reduced to the form

$$\Psi(X) = \frac{1}{(2\pi)^{n/2}} \int \ldots \int \exp[i\Phi(s; X)] d^n s, \tag{1.29}$$

where s represents the state variables and X the control parameters (for the case of the rainbow there is only one of each, so $n = 1$). These integrals stably represent the wave patterns near caustics. The *corank* of the catastrophe is equal to n: it is the minimum number of state variables necessary for Φ to reproduce the stationary-point structure of ϕ; the *codimension* is the dimensionality of the control space minus the dimensionality of the singularity itself. It is interesting to note that in ray catastrophe optics, the state variables s are removed by differentiation (the vanishing of the gradient map); in wave catastrophe optics they are removed by integration (via the diffraction functions). For future reference we state the functions $\Phi(s; X)$ for both the fold (A_2) and the cusp (A_3) catastrophes; the list for the remaining five elementary catastrophes can be found in the references above. For the fold

$$\Phi(s; X) = \frac{1}{3}s^3 + Xs, \tag{1.30}$$

and for the cusp

$$\Phi(s; X) = \frac{1}{4}s^4 + \frac{1}{2}X_2 s^2 + X_1 s.$$

By substituting the cubic term (1.30) into the integral (1.29), it follows that

$$\Psi(X) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} \exp\left[i\left(\frac{s^3}{3} + Xs\right)\right] ds = (2\pi)^{1/2}\,\mathrm{Ai}(X),$$

where the integral here denoted by $\mathrm{Ai}(X)$ is one form of the Airy integral. It will be encountered in several different guises as we proceed through the book. For $X < 0$ (corresponding to $\theta > \theta_R$) there are two rays (stationary points of the integrand) whose interference causes oscillations in $\Psi(X)$; for $X > 0$ there is one (complex) ray that decays to zero monotomically (and faster than exponentially). This describes diffraction near a fold caustic. In 1838 the Astronomer Royal Sir George Biddle Airy introduced this function to study diffraction along the asymptote of a caustic (although he did not express it in these terms) and provided a fundamental description of the supernumerary bows [23]. (There is also a sequel to this paper, published ten years later [268], which contains a transcript of a fascinating letter from the British mathematician Augustus De Morgan (1806–1871)). As noted above, this integral (in one form or another) will be a recurring topic in several later chapters. The corresponding integral for the cusp catastrophe is frequently referred to as the *Pearcey integral* [157].