MAKING LIGHT OF MATHEMATICS

MICHAEL BERRY

ABSTRACT. Summary of AMS Gibbs Lecture, delivered at San Diego, CA, 6 January 2002.

1. Introduction

According to the J. W. Gibbs Fan Club Homepage, Gibbs “was a quiet, dignified man whose life was presumably boring (part of the reason why no one has heard of him)”, and “a physicist, chemist, and mathematician—a sort of all-around hard sciences guy.” Gibbs was a mathematical physicist, so at least the second statement is correct.

Nowhere are the intimate connections between mathematics and physics more immediately apparent than in optics; with our own eyes, we can see through physical phenomena almost directly to the conceptual structures underlying them. Risking the wrath of philosophers, I use the term mathematical phenomena to describe these structures. In the 2002 Gibbs Lecture, I described some mathematical phenomena and their optical counterparts—turning mathematics into light, as it were. The lecture was based almost entirely on images—photographic, computer-simulated and demonstrated live—and is impossible to reproduce in print. What follows is a summary with references, organised by the mathematical phenomena.

2. 1 + 1 = 2, 1 + 1 ≠ 2

Two candles, or two flashlights, are twice as bright as one. Knowledge that light is waves makes this problematic, because the addition of two pure lights with unit amplitudes and phases 1 and 2 gives the intensity

(1) intensity = \left| \exp(i\phi_1) + \exp(i\phi_2) \right|^2 = 2 + 2 \cos(\phi_1 - \phi_2),

and the interference term, involving the cosine, makes the intensity vary between 0 and 4. But flashlight is messy light, not pure light, and the average over messiness (random fluctuations in the phases, from nonmonochromaticity, finite resolution of detectors, etc.) suppresses the interference, restoring 1 + 1 = 2.

Nowadays, the suppression of interference by averaging is called decoherence. The idea might seem trivial but is implicated in the emergence of the classical world from the quantum world as a result of uncontrolled influences from the environment (that is, from outside the system being studied).
3. Stable singularities of gradient maps

The bright points of reflected sunlight on the surface of the sea are images of the many-to-one gradient map defining light rays from the sun to the eye via the sea \[53\], \[54\], \[55\]. The map is

\[
\nabla_x \Phi(x; C) = 0 \quad (x = \{x_i\}, C = \{C_n\}),
\]

where \(\Phi\) is the length of a light path, depending on variables \(x\) (position on the water surface) and parameters \(C\) (position of the eye, time...).

Singularities of the map, satisfying

\[
\det \left( \frac{\partial^2 \Phi}{\partial x_i \partial x_j} \right) = 0
\]
as well as (2), correspond to the coalescence of images, where a caustic (focal) surface in space crosses the eye. The rapid occurrence of many such events gives the sea its sparkling appearance. The rainbow \[52\] exemplifies the simplest caustic, where two rays (deflected by refraction and reflection in a raindrop) coalesce (in direction space). ‘Natural focusing’ \[14\], involving caustics that are stable (under diffeomorphism), rather than the symmetry-dominated artificial focusing required for the lenses and mirrors of optical instruments, is described by the singularities classified in catastrophe theory \[2\], \[33\], \[65\].

Catastrophes are important in the physics of waves as well as rays, because light is brightest near caustics and spacings between interference fringes shrink slowest as the wavelength gets smaller. Both effects are captured by a wavelength scaling law \[77\], \[9\] satisfied by the wave \(\Psi\) corresponding to parameters \(C\) with light of wavelength \(\lambda = 2\pi/k\), when \(k\) is large (short-wave asymptotics). The scaling law is

\[
\Psi(k, C) = k^{\beta} \Psi(1; \{k^{\gamma_j}C_j\}),
\]

where the exponent \(\beta\) describes the divergence of intensity \(|\Psi|^2\) as \(k \to \infty\) (as \(k^{2\beta}\)), and the exponents \(\gamma_j\) describe the shrinking fringe spacings in the different directions in parameter space (as \(k^{-\gamma_j}\)). \(\Psi\) denotes the ‘diffraction catastrophe’ integral

\[
\Psi(1; C) = \prod_j \int_{-\infty}^{\infty} dx_j \exp\{i\Phi(x; C)\},
\]

where \(\Phi\) is a polynomial normal form from the list of stable singularities \[4\], \[5\], \[3\].

Diffraction catastrophes constitute a class of special functions \[21\], outside the conventional hypergeometric class, describing the intricate wave patterns that decorate geometrical caustics. The first two, corresponding to diffraction near the fold (rainbow) and cusp catastrophes, have been known for a long time \[1\], \[64\], but the class as a whole is new. Mathematical properties of the diffraction catastrophes form a chapter in the forthcoming Digital Library of Mathematical Functions \[39\]. Figure \[1\] shows several representations of the cusp diffraction catastrophe.

4. Divergent series

Two rays reach each direction on the bright side of a rainbow, and their interference is visible as supernumerary rainbows \[52\], \[8\] decorating the main arc under conditions where all the raindrops have similar sizes. The two rays coalesce at the
rainbow angle, and on the dark side there are no real rays. But there are complex images of the gradient map, corresponding to exponentially decaying (evanescent) waves. So, far from the caustic, the rainbow light is represented by two (complex, interfering) exponentials (bright side) and one (real, decaying) exponential (dark side).

The two regimes (bright and dark) can be connected by a path in complex direction space that avoids the caustic. The transformation between different numbers of exponentials, representing the same function, is Stokes’s phenomenon [74], [73]. In its simplest form (as in the rainbow) it occurs on Stokes lines, where one exponential changes its coefficient (e.g. appearing or disappearing) when maximally dominated by the other. Stokes’s phenomenon originates in the divergences of the asymptotic series (in powers of $1/k$) that multiply each of the exponentials representing the function in different regions. The switching on and off of the small exponentials is described in a very wide variety of cases by an error function [20], [19] obtained by resumming the divergent tails of the series [38], [41], [40].

Caustic singularities describe the violent births and deaths of real rays; Stokes’s phenomenon is a different type of singularity (nonlocal bifurcation) describing the gentle births of complex rays—the quietly beating heart of asymptotics.

5. **Zero**

At a zero of the complex wave representing light (Figure 1b), the phase is singular (Figure 1c). Phase singularities are codimension-2 phenomena, representing perfect destructive interference ($1 + 1 = 0$). These dark singularities of waves [61], [60], [71], [70] are complementary to the bright caustic singularities of rays. They are generic features of waves of all types [18], e.g. sound [79], and occur not only in diffraction catastrophes [30] but also in the random waves [25] that can represent optical speckle, black-body radiation, or the electron waves in a quantum dot.

Phase singularities have recently been extensively studied in optical experiments [69], [78]. Interest in them, (‘much ado about nothing’ [13]) reflects a contemporary preoccupation with zero [7], [46], [66].

6. **Knots**

In three dimensions, phase singularities are lines (threads of black light) that can be curved and form closed loops. They can also be knotted and linked, as shown by explicit prescriptions for generating torus knots in optical beams [23]. An analogous construction shows that knotted zeros can be created in quantum states, e.g. in the hydrogen atom [11].

As a parameter (e.g. time) varies, knots can be created and destroyed by a stable process [24] in which two strands manoeuvre into local parallelism and then switch their (locally hyperbolic) branches.

7. **Gauss sums**

The wave representing light at a point P beyond a coherently illuminated diffraction grating with $N$ narrow slits is given by the irrational incomplete Gauss sum

\[ S_N(\tau) = \sum_{n=1}^{N-1} \exp(i\pi \tau n^2), \]
where \( \tau \) depends on \( P \), the slit spacing \( a \) and the wavelength \( \lambda \). This sum of unit vectors in the complex wave-amplitude plane traces elaborate patterns of ‘curlicues’ as \( N \) increases, generalizing the Cornu spiral that results by replacing the sum in (6) by an integral.

Curlicues are an ‘arithmetic microscope’, revealing more of the arithmetic nature of \( \tau \) as \( N \) increases. The sum (6) satisfies an asymptotic \(( N \gg 1)\) renormalization (coarsening) transformation [43], [26], namely

\[
S_N(\tau) \approx \exp\left(\frac{\sqrt{\frac{a}{\lambda}}}{\pi} K^{1+\text{int}(\tau^{-1})} S_{\text{int}(N\tau)}(\tau_1(\tau))\right),
\]

where \( K \) denotes the operation of complex conjugation and

\[
\tau_1(\tau) = \tau^{-1}(\text{mod } 1)(\text{int}(\tau^{-1} \text{ even})
\]

\[
= 1 - \tau^{-1}(\text{mod } 1)(\text{int}(\tau^{-1} \text{ odd})
\]

Repeated application of (7) (restricting \( \tau \) to the interval \((0 \leq \tau \leq 1)\) without loss of generality) leads to a comprehensive understanding of the hierarchical structure of the curlicues for different classes of number \( \tau \) by relating the patterns that would be seen under different magnifications. Quadratic irrationals correspond to fixed points of the map (8) or a power of this map. The map is ergodic and mixing, so the pattern for any typical \( \tau \) is related by scaling to the patterns for all other typical \( \tau \).

For a diffraction grating with infinitely many slits, patterns repeat at observation distances \( z \) that are integer multiples of \( z_T = a^2/\lambda \). This is the Talbot effect [75], [63]. For fractional distances \( z = (p/q)z_T \), the patterns [27] consist of \( q \) coherently superposed images of the grating. If this is described by its transparency \( g(x) \) (with \( g(x+1) = g(x) \)), the fractional Talbot images are

\[
\psi(x, p/q z_T) = \frac{1}{\sqrt{q}} \sum_{s=1}^{q} g\left(x - a \frac{s}{q} - \frac{1}{2}q e(p)\right) \exp\{i\phi(s, q, p)\},
\]

where \( \phi \) is the phase of the rational Gauss sum

\[
\exp\{i\phi(s, q, p)\} = \frac{1}{\sqrt{q}} \sum_{n=1}^{q} \exp\left\{i\frac{\pi}{q} \left(-n^2p + n \left(s + \frac{1}{2}q e(p)\right)\right)\right\}.
\]

If propagation distance \( z \) is replaced by time \( t \), the parabolic wave equation describing diffraction becomes the Schrödinger equation governing the evolution of quantum wavefunctions. The analogue of the Talbot effect is then the phenomenon of quantum revivals [6], [17], in which an initially compact wavepacket (representing, for example, an electron in an atom) first spreads and later reconstitutes itself (or, in the fractional case, several copies of itself). Revivals have been observed [80].

8. Fractals

For observation at irrational distances \( z/z_T \), the patterns (9) are formed by the superposition of infinitely many images of the grating. In the familiar Ronchi grating (equal transparent and opaque bars, i.e. \( g(x) = \Theta(\cos(2\pi x/a)) \), where \( \Theta \) is the unit step), the resulting intensity graph is a fractal curve with dimension \( D = 3/2 \) [27]. The fractal is anisotropic: observation at constant \( x \) gives an intensity
graph (as a function of $z$) that is a fractal with $D = 7/4$, and there are diagonal lanes in $xz$ space where the graph has $D = 5/4$.

In the quantum analogue [17], the counterparts of these transverse, longitudinal and diagonal optical fractals are space fractals, time fractals and spacetime fractals for any initial wavepacket with a discontinuity (e.g. the particle in a one-dimensional box with Dirichlet conditions at the walls and initial state $\psi = 1$). For both the optical and quantum cases, pictures of the full wave in the $xz$ or $xt$ planes resemble rich ‘carpets’ [29], woven by interference under circumstances of extreme coherence.

A different type of fractal wave occurs in unstable lasers [68]. In the usual laser cavities, the modes, confined by two concave mirrors, are narrow gaussian beams. In unstable lasers, one or both of the mirrors is reversed, so the rays are defocused rather than focused and spread to fill the cavity. With a transverse coordinate $x$, measured in units of the half-width of the smaller mirror, modes $u(x)$ are eigenfunctions of the round-trip wave map $T$

$$T u(x) = \sqrt{\frac{A}{2\pi M}} \int_{-1}^{1} dy \exp \left\{ \frac{1}{2} iA \left( y - \frac{x}{M} \right)^2 \right\} u(y) = \gamma u(x). \tag{11}$$

Here $M$ ($>1$) is the magnification, $A$ is the ‘Fresnel number’ characterising the laser and inversely proportional to $\lambda$, and $\gamma$ is the mode eigenvalue.

The integral operator $T$ in (11) is nonunitary because some light misses the mirrors and gets lost at each reflection; therefore $|\gamma| < 1$. For the lowest-loss mode (the one with $|\gamma|$ closest to unity), $|\gamma| \approx 1/\sqrt{M}$. The fine structure of the mode is determined by multiple reflections of edge-diffracted waves [45], [72], leading to a representation for $u(x)$ as a truncated analogue of the series for the Weierstrass nondifferentiable function [28]. The largest $x$-scale of the mode is unity (size of the smaller mirror), and the smallest scale is $1/A$, which vanishes in the short-wave limit.

The resulting intensity graph of the lowest-loss mode is an asymptotically emergent fractal [56], [10], [47], [37], [59] with $D = 2$ [32], [10] that is nondifferentiable and with the last vestige of continuity. For the more lossy modes, $D$ is smaller and given (G. New, private communication) by

$$D = 2 - \frac{\log |\sqrt{M}/\gamma|}{\log M}. \tag{12}$$

9. Matrix degeneracies

For a nonunitary operator, such as $T$ in (11), the right eigenvector $|u\rangle$ (eigenvector of $T$) is different from the left eigenvector $|\nu\rangle$ (eigenvector of Hermitian conjugate of $T$). This has the (undesirable) physical consequence that the linewidth of an unstable laser exceeds the quantum-optics minimum by the Petermann excess-noise factor [58]

$$K = \frac{1}{|\langle \nu | u \rangle|^2}. \tag{13}$$

For any operator, $K \geq 1$, with equality for unitary operators. Computations [48], [57] reveal that $K$ can take very large values, in the form of peaks as a function of the parameter $A$ in (11).
This phenomenon is a consequence of the radically different degeneracy structure possessed by nonunitary operators in comparison with unitary (or Hermitian) ones. Approaching a degeneracy, $|u\rangle$ and $|\nu\rangle$ become orthogonal (rather than parallel as for unitary operators), so $K \to \infty$. Degeneracies of general complex matrices have (real) codimension 2, and so do not usually occur for real $A$. The observed peaks are associated with complex degeneracies near the real $A$ axis [12]: they are resonances of a peculiar square-root-of-Lorentzian form whose asymptotics can be comprehensively understood in terms of the edge-wave representation of $u(x)$—another example of ‘real physics in the complex plane’ [15], [31].

Quite different are the degeneracies of real symmetric matrices, though these also have codimension 2. Such matrices govern the propagation of polarized light in transparent anisotropic materials [35]. These $2 \times 2$ matrices, whose eigenvectors give the two orthogonal polarizations that can propagate without change and whose eigenvalues give the corresponding refractive indices, depend on the (2-parameter) direction $\mathbf{s}$ of the light. For a generic (biaxial) material, degeneracies occur for four values of $\mathbf{s}$, namely the optic axes. Near the optic axes, the eigensurfaces are locally double cones, connected ‘diabolically’ [34] at the optic axes.

A readily available biaxial material enabling these phenomena to be demonstrated easily and beautifully is overhead-projector transparency foil. Make this the ‘filling’ of a ‘black sandwich’ whose ‘bread’ is two crossed polarizing sheets (i.e. with their optic axes perpendicular). Then view any bright surface of diffuse light (e.g., the sky) through the sandwich. The light paints a ‘bullseye’ of colored interference rings crossed by a black brush [22]. The rings are contours of constant eigenvalue separation, with the degeneracy at the centre, and the brush is a consequence of the sign reversal [44], [76] of each eigenvector when smoothly continued round each optic axis—an elementary example of the geometric phase [16], [67].

If the transparent material is chiral as well as anisotropic, the governing matrix is not real symmetric but complex Hermitian; this was studied by Gibbs [22].

REFERENCES


[64] T. Pearcey, The structure of an electromagnetic field in the neighbourhood of a cusp of a caustic, Phil. Mag. 37 (1946), 311–317. MR 8:605d

H H WILLS PHYSICS LABORATORY, TYNDALL AVENUE, BRISTOL BS8 1TL, UNITED KINGDOM
URL: http://www.phy.bris.ac.uk/staff/berry_mv.html
Figure 1. The cusp diffraction catastrophe. (a) Photograph of a screen illuminated by a green laser beam that has passed through smoothly undulating bathroom window glass whose irregularities are about 1mm across. (b) Intensity $|\Psi|^2$ of the (codimension 2) cusp diffraction catastrophe (Pearcey integral), colour-coded as in the inset (red: maximum; black: zero). (c) Phase arg $\Psi$ of the Pearcey integral, colour-coded by hue, with the zeros of $\Psi$ revealed as phase singularities where all colours meet.