Topological Singularities
in Wave Fields

Mark Richard Dennis

H. H. Wills Physics Laboratory
University of Bristol

A thesis submitted to the University of Bristol in
accordance with the requirements of the degree of
Ph.D. in the Faculty of Science

November 2001

Abstract

This thesis is a study of the natural geometric structures, arising through interference, in fields of complex waves (scalars, vectors or tensors), where certain parameters describing the wave are singular. In scalar waves, these are phase singularities (also called wave dislocations), which are also nodes (zeros of amplitude): in two dimensional fields they are points, and in three dimensions, lines. The morphology of dislocation points and lines is studied in detail, and averages of their geometrical properties (such as density, speed, curvature and twistedness) are calculated analytically for isotropically random gaussian ensembles (superpositions of plane waves equidistributed in direction, but with random phases). It is also shown how dislocation lines may be knotted and linked, and a construction of torus knots in monochromatic waves is studied in detail, using experimentally realisable beams. In vector waves, the appropriate fields are described geometrically by an ellipse at each point (the polarization ellipse). Their singularities, occurring along lines in three dimensions, are where the ellipse is circular (C lines) and linear (L lines); in two dimensional fields, possibly representing the transverse plane of paraxial polarized light waves, there are C points, but still L lines. The geometry of these singularities is considered, and analytical calculations for their densities in isotropic gaussian random vector waves are performed. The C and L singularity structures are generalised to fields of spinors using the Majorana sphere (vector fields have spin 1), and singularities in rank two tensor waves (spin 2) are briefly discussed.

For my parents

Acknowledgements

First and foremost, I would like to express my gratitude to my supervisor, Michael Berry, for suggesting this Ph.D. topic in the first place, and for his patience helping me along the way. My appreciation of the beauty of nature and the role of physics in it has been completely transformed from what he has taught me. I owe equally much to the insight of John Hannay, with whom I have enjoyed many useful lunchtime discussions. I thank John Nye for his wisdom and encouragement. It has been a privilege to play a minor part in the unfolding of Bristol singularity physics.

This work is also a testament to the breadth and friendliness of the Bristol Physics Theory Group: there isn't one of the staff here who I have not approached at some stage on some aspect of this research. I have also benefitted from discussions with Arnd Bäcker, Karl-Frederick Berggren, Johannes Courtial, Isaac Freund, Miles Padgett, Jonathan Robbins, Marat Soskin, Michael Vasnetsov and Art Winfree. This work would not have been possible without a University of Bristol Postgraduate Scholarship.

For helping me survive the last three years, I am indebted to my fellow monkeys Duncan, Kevin, Jon, Muataz, Jason, Jorge, Markus, Matt, Ben, Dan, Denzil, Andy and Andy. Joe Brader deserves special mention; it will be difficult to distract him as easily now our desks are separated by the Atlantic. I have enjoyed my heated and wide ranging arguments with Paul Upton.

Outside the Physics Department, I would like to thank Kevin Duggan and the Bristol Chamber Choir for providing much Wednesday evening relief and amusement. Mark put up with me for three years, and providing the entertainment facilities in our shared accommodations : *"Squeaky wheel gets the kick!"* Michael and Monica Berry's generosity and hospitality have helped me throughout, especially in the last awkward months.

Final words of thanks go to Caroline, whose love and patience have been a constant reminder that there is more to life than work. Over the three years she has always given me a refuge, whether to the north, east, west or south of Bristol. Of course, my parents deserve the last word; where would I be without them?

Author's Declaration

I declare that the work in this thesis was carried out in accordance with the Regulations of the University of Bristol, between October 1998 and October 2001. The work is original except where indicated and no part of the thesis has been submitted for any other degree. A number of original results presented here were found in collaboration with my supervisor, Professor Michael Berry. Any views expressed in the thesis are those of the author and in no way represent those of the University of Bristol. The dissertation has not been presented to any other University for examination either in the United Kingdom or overseas.

'To listen to one indefatigable lichenologist commenting upon the work of another indefatigable lichenologist, such things force one to realise the unfaltering littleness of man.'

H.G. Wells *The Food of the Gods*, in *The scientific romances of H.G. Wells*, London, Victor Gollancz Ltd, 1933

Contents

List of Figures

List of Tables

Chapter 1

Introduction: What is a singularity?

'Our task is to find in all these factors and data, the absolute, the universally valid, the invariant, that is hidden within them.'

Max Planck, *Scientific Autobiography and Other Papers*, New York: Philosophical Library, 1949

1.1   General Introduction

Many phenomena in nature, such as light, sound, thermal radiation and quantum mechanical matter, can be described by waves. This thesis is a study of wave interference, that is to say, what happens when the wave disturbance from different sources adds together (*constructive interference*) or cancels out (*destructive interference*). Across an entire field of interfering waves, there are places where either of these types of interference can happen. Thomas Young, in November 1801, announced his discovery of interference in a light beam, thereby proving that light physically is a wave [You02]. For fields of interfering waves, in three dimensions, the destructive interference occurs along lines, resulting in threads of darkness (silence, etc), and in two dimensions, at points. These interference structures were discovered as a general phenomenon of wave physics by Nye and Berry in 1974, where, in analogy with the defects of crystal lattices, they were called wave dislocations.

Wave dislocations are not objects in the usual sense (like atoms): they do not have an independent existence, but are specific features of the patterns set up by the interfering waves. Such phenomena are called *morphologies*, another example of which is a shadow (this a morphology of light rays, not waves). The importance of wave dislocations is brought out when one asks what physical quantities describe the wave disturbance at each point: in addition to the size of the wave (the amplitude, a nonnegative real number), there is also a *phase*, that is, the angle which determines the point of the wave cycle (which changes, for example, as time evolves). Where the amplitude of the wave is zero, the phase cannot be determined (since the wave is zero for all phases): the wavefield zeros are also *phase singularities*, and are the most significant features of the phase landscape in the field. A simple example of a phase singularity (not in a wave) is the singular time zone at the north pole (where the phase angle corresponds to the position of the hour hand on a watch, which may validly point to any hour at the north pole).

It is remarkable that, if any wavefield is chosen at random (out of an appropriate ensemble), these singularities occur naturally throughout the field, out of the random interference pattern, and part of the work described here is an exact mathematical calculation of the densities of dislocations in general kinds of random wavefield, as well as the statistical distributions of geometric properties such as curvature, speed (if they are moving) and twistedness. These calculations apply to the threads of silence in a noisy room, or the threads of darkness from light emitted from a thermal radiator (ie a black body).

However, we may also choose to manipulate waves (for instance light in a laser beam) in order to configure the dislocations in the wave into desired forms; we also describe a method for creating dislocation loops that are knotted or linked in physically realisable beams. This construction takes advantage of the mathematical structure of the wave around the singularity, and the interference pattern near the knot has a detailed and subtle structure.

Not all waves are just described by amplitude and phase; they may also have *polarization*, where the wave disturbance occurs in a certain direction or directions (examples include (un)polarized light or elastic waves in a solid). In this case, there are too many variables for the wavefield naturally to vanish on lines, but there are other types of singularity in these wavefields, discovered and measured by Nye and Hajnal [Nye83b, Nye83a, NH87, Haj87a, Haj87b, Haj85]. The mathematical formalism of these

singularities is reviewed here, and we present calculations of their statistical densities in random vector waves. The singular polarization structures for vectors may be generalised to more complicated types of wave (tensor waves, with higher spin), and an outline of how this happens is presented in the final chapter.

Although dislocations and other wave singularities have been known and studied for a long time, this new work is original in several ways. The initial motivation was purely statistical, to generalise the methods used in, and types of quantity considered by, the statistical analyses of [Ber78, Fre94], leading to the publication of three articles in research journals, [BD00], [BD01c], [Den01b] as well as a conference proceedings article, [Den01a]. However, more general investigation of the topology of dislocation loops led to the rediscovery of the 'twisted loop' theorem (originally discovered and investigated by Winfree and coworkers), and thus the possibility of experimentally realisable dislocation knots, as described in [BD01a] and [BD01b]. The generalization of polarization singularities to waves of higher spin is not yet complete, and has not been published at this time.

1.2 What is a singularity?

In this section we shall explore the simple but fundamental notions associated with phase singularities, as well as the related singularities of real vector fields and line (ellipse) fields in two dimensions. These mathematical structures occur in other contexts than wave interference, and the discussion here is general.

Phase singularities occur at the zeros (nodes) of complex scalar fields, that is functions from space (of either two or three dimensions) to the complex numbers (analytic when convenient). Unless otherwise stated, this function shall always be represented by ψ, so $\psi : \mathbb{R}^2, \mathbb{R}^3 \longrightarrow \mathbb{C}$.

Sometimes, ψ is time dependent as well as space dependent. Where confusion will not ensue, explicit functional dependence on space or time will be suppressed (ie $\psi = \psi(\mathbf{r}), \psi(\mathbf{r}, t)$).

The complex field ψ can be written in terms of its modulus ρ and argument χ or its real and imaginary parts, ξ, η :

$$\psi = \rho \exp(\mathrm{i}\chi) = \xi + \mathrm{i}\eta, \qquad (1.2.1)$$

where $\rho > 0, \chi, \xi, \eta$ are real, and the angle χ is singlevalued modulo 2π. Since ψ is to

denote a wave, the modulus ρ is the wave *amplitude*, and the argument χ is the *phase* of the wave, and plays a crucial role in what follows.

The nodes of ψ are the places $\mathbf{r}$ in space for which

$$\psi(\mathbf{r}) = 0 \qquad (1.2.2)$$

and, usually, are points in two dimensions, and lines in three. What is meant by 'usually' (ie generically) will be discussed in the next section. For now, we restrict attention to the geometry of nodal points in two dimensions, with points labelled $\mathbf{R} \equiv (x, y)$ in cartesian coordinates, (R, ϕ) in plane polars (see section 1.9 for a summary of notational conventions).

The simplest function with a phase singularity is simply the natural map from cartesian space to the complex plane,

$$\psi(x, y) = x + \mathrm{i}y = R \exp(\mathrm{i}\phi) \qquad (1.2.3)$$

which is zero at the origin. The phase here is the polar angle ϕ, which is defined everywhere except the origin, where R is 0. ϕ, as a function of position in the plane, is not continuous; following any smooth line of constant ϕ through the origin causes a jump of π. There is therefore no way of ascribing to ϕ a value at the origin: the phase of ψ is singular. It is a necessary condition that R be zero at the singularity, otherwise the argument of the nonzero ψ in (1.2.3) would not be defined. Phase singularities and zeros are therefore equivalent, and the two terms shall be used interchangeably.

A reason that phase singularities are important is that their presence determines the phase structure around them. In particular, consider the line integral

$$s = \frac{1}{2\pi} \oint_C \mathrm{d}\chi = \frac{1}{2\pi} \oint_C \nabla\chi \cdot \mathrm{d}\mathbf{R}, \qquad (1.2.4)$$

where C is some closed nonselfintersecting loop directed in the positive (anticlockwise) sense in the plane, not passing through any node (so $\nabla\chi$ is well-defined). Since χ is singlevalued modulo 2π, s is an integer (possibly positive, negative, or zero). Topologically, the image $\chi(C)$ of the path C is a closed loop in the space of angles, winding round s times (positive if χ increases around the loop, negative if it decreases). s is therefore referred to as the *winding number* of the loop. Now, if C encloses the origin in equation (1.2.3), $s = 1$, and $s = 0$ otherwise. If ψ in (1.2.3) were replaced by its conjugate $\psi^* = x - \mathrm{i}y = R \exp(-\mathrm{i}\phi)$, then $s = -1$, since $\chi = -\phi$ for this function. Similarly,

for positive integer n, any loop around 0 of the function ψ^n has winding number n (and similarly, $(\psi^*)^n$ has $-n$). This topological number, being shared by all loops enclosing the origin in (1.2.3) (and, more generally, any nodal point) is thus a property of the phase singularity, called its *topological charge* [Hal81, Ber98], and, in the case of wave dislocations (when ψ represents a wavefield), the *dislocation strength.*

A simple dislocated wave in two dimensions that has often been discussed (for example, by [NB74, Ber81, Nye99]) has the equation

$$\psi = (x + \mathrm{i}y)\exp(\mathrm{i}ky) \qquad (1.2.5)$$

(actually, it is an approximation to a solution of the Helmholtz equation (1.5.2)). The dislocation (as phase singularities in waves are called) at the origin has strength 1, regardless of the sign of the wavenumber k, and locally to the dislocation (on a scale where $ky \approx 0$), equation (1.2.5) reduces to (1.2.3). The contours of constant phase (the *wavefronts*) emerge uniformly from the singularity at 0, and, further away, arrange into parallel lines normal to the y direction, owing to the plane wave factor $\exp(\mathrm{i}ky)$. The pattern of wavefronts near the dislocation is shown in figure (1.1). This wavefront structure is the origin of the term dislocation for nodes in wavefields; Nye and Berry ([NB74]) coined the term since figure (1.1) has a structure similar to the atomic planes in a crystal near to an edge dislocation.

As mentioned above, across the singularity, the phase changes by π. The zero contours of the real part ξ are labelled by one of two phases, 0 and π, which alternate at each node; the position of the node cannot be found by inspection of the real part alone (only that the node lies along the zero contour), and the phase (out of 0 and π) similarly cannot be determined. Finding the node using the imaginary part alone has similar problems (with the two phases being $\pi/2, 3\pi/2$) but the positions of the singularities can be determined as the intersections of the two sets of zero contours of real and imaginary parts. The behaviour of the other phase contours are easily found from ξ and η. Phase contours, wavefronts and phase topology are discussed in more detail in section 2.1.

Phase singularities occur very generally whenever there is an angle continuously dependent on two or three spatial parameters. One example, frequently taken advantage of in representation of phase fields (although not here) is the so-called colour wheel: we see the various hues (red, orange, yellow, etc) as a continuum, where purple appears with the other spectral colours (although it is a mixture of red and blue), joining the two ends of
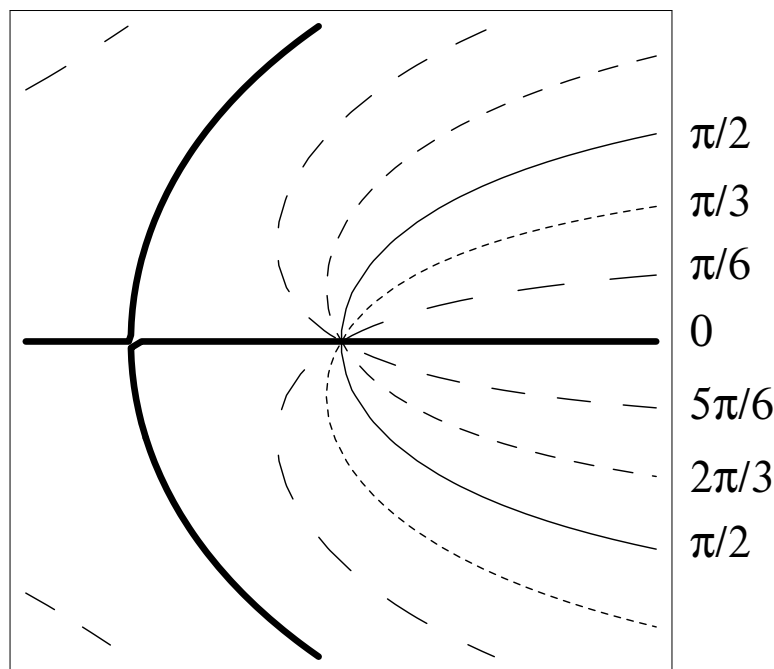
Figure 1.1: Wavefronts (contours of constant phase, mod π) of the edge dislocation (1.2.5), scaled such that $k = 1$.

the visible spectrum into a circle. The hues may be continued towards the centre of the circle, but no colour can be put at the centre which continuously joins up with the others (it can only be grey). The hue representation of phase has been used to pick out the phase behaviour near singularities by Winfree [Win87].

Another example of a phase singularity, mentioned already, is the problem of the correct time zone at the north pole. Consider an idealised globe, where the time zones are separated by geodesic lines, independent of territories or borders, as in figure (1.2). The north and south poles lie at points where all these lines intersect, so are not in any unique zone: it is a phase singularity of the position of the hour hand on a watch (of course, the variation of local time on the earth ought to be continuous with respect to east-west position, but for convenience is discretised into one hour jumps). Fortunately, this does not often give rise to practical problems, because the population at the poles is rather sparse.[1]

---

[1] As J.F. Nye informs me, the presence of the British Antarctic survey at the south pole has led to Greenwich Mean Time taken as standard at the south pole, and compass directions are taken with respect to the Greenwich meridian.
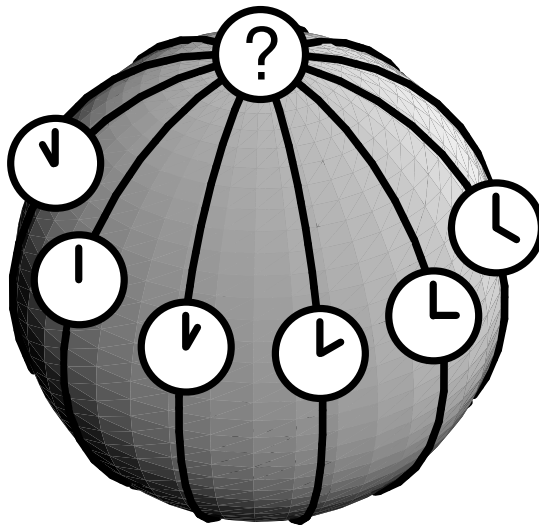
Figure 1.2: The phase singularity of time at the north pole (with idealised time zones).

The topological charge of the polar singularities (north pole +1, south pole −1) have interesting implications for a closed loop enclosing the singularity, that is, a closed E-W journey around the world. Without the international date line, continuous adjustment to local time gives rise to the acquisition (or loss) of a day on return to the starting point. This surprising idea was the punchline of Jules Verne's *Around the world in 80 days*, for, as Phileas Fogg discovered, he

> ... had, without suspecting it, gained one day on his journey, and this merely because he travelled constantly *eastward;* he would, on the contrary, have lost a day, had he gone in the opposite direction - that is, *westward.*
>
> In journeying eastward he had gone towards the sun, and the days therefore diminished for him as many times four minutes as he crossed degrees in that direction ... three hundred and sixty degrees, multiplied by four minutes, gives precisely twenty-four hours - that is, the day unconsciously gained.

This problem is avoided with the international date line, which introduces a discontinuity of a day's jump along a line joining the two poles. This line solves the problem of multivaluedness of the date (at the cost of introducing a 2π discontinuity), but its nature is easy to misunderstand (see, for example, the confusion of the protagonist in Umberto Eco's *The Island of the Day Before*).

It is no coincidence that the number of phase singularities on the globe is equal to the Euler characteristic of the sphere, which, by the Poincaré-Hopf theorem [Mil65], is the total Poincaré index of any smooth vector field on that surface. There are several connections between phase singularities in complex scalar fields and real vector field singularities, as we now describe.

Consider a vector field in the plane $\mathbf{V}(\mathbf{R})$, with components

$$\mathbf{V} = (V_x, V_y)\,\text{(cartesian)} = (V, \theta)\,\text{(polar)}, \tag{1.2.6}$$

such that V_x, V_y are smooth. V and θ are not usually smooth, and in fact θ is singular when V is zero (for example the origin when $\mathbf{V} = \mathbf{R}$). Around any closed nonselfintersecting loop C (avoiding places where $\mathbf{V} = \mathbf{0}$), take a line integral of θ, analogous to equation (1.2.4):

$$I_P = \frac{1}{2\pi} \oint_C \mathrm{d}\theta = \frac{1}{2\pi} \oint_C \nabla\theta \cdot \mathrm{d}\mathbf{R}. \tag{1.2.7}$$

As with phase χ before, θ is well-defined everywhere that $\mathbf{V}$ does not vanish, so the integral is some (positive, negative or zero) multiple of 2π, and I_P is an integer. C can be deformed continuously without changing I_P (which can only change by integer jumps), provided it does not cross a zero. On taking a very small path around a point zero of $\mathbf{V}$, one finds that this integer is usually nonzero, and is called the *Poincaré index* of the zero. The singularity in θ is sometimes called a direction field singularity, and is truly a singularity of the unit vector field $\mathbf{V}/V$.

There are three distinct types of singularity with Poincaré index ± 1, whose linear behaviour around the zero (taken to be at the origin) is given by a matrix $\mathbf{M}$, where in a neighbourhood of the origin,

$$\mathbf{V} = \mathbf{M}\,\mathbf{R}. \tag{1.2.8}$$

It is easy to see that the Poincaré index, defined by equation (1.2.7), is equal to the sign of the determinant of $\mathbf{M}$.

The first type of vector field singularity is a *circulation*, given by the matrix (up to an unimportant nonsingular linear transformation)

$$\mathbf{M}_{\mathrm{circ}} = \pm \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \tag{1.2.9}$$

where the sign determines the sense of the circulation around the origin. It has determinant 1, trace 0 and eigenvalues $\pm i$. The second type (again two fields, since the vectors can point in either direction) are *sources* and *sinks*, with matrix (up to transformation)

$$\mathbf{M}_{\mathrm{so/si}} = \pm \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \tag{1.2.10}$$

where sources have the $+1$ prefactor, sinks -1. They have Poincaré index 1, trace ± 2 and repeated eigenvalues 1 or -1. The final type, a *saddle* (or saddle point) has the matrix (up to transformation)

$$\mathbf{M}_{\mathrm{sad}} = \pm \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{1.2.11}$$

Taking the negative of $\mathbf{M}_{\mathrm{sad}}$ is equivalent to a rotation by $\pi/2$ or a reflection in the line $y = x$. It has index -1, trace zero and eigenvalues ± 1. After an affine transformation, the determinant of any of the $\mathbf{M}$ may change, but its sign does not. The circulations remain traceless (although the circular flow lines near the singularity may be deformed),

(a)   (b)   (c)
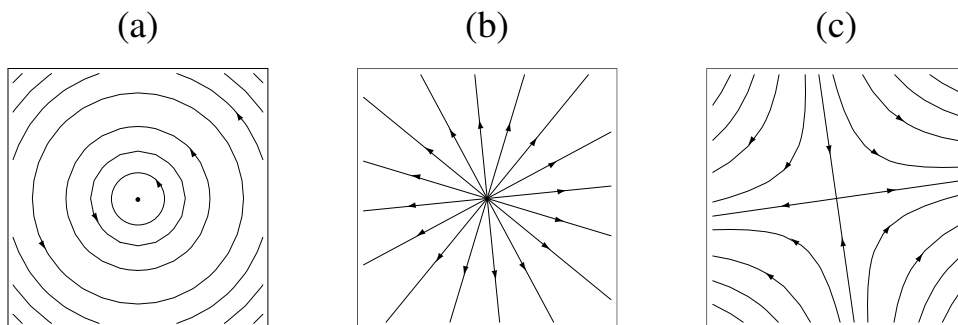
Figure 1.3: The three types of planar vector field singularity with index ± 1 : (a) circulations; (b) source and sink; (c) saddle.

but the saddle may acquire a nonzero trace. These three types of singularity are the only three with unit Poincaré index (circulations, sources, sinks $+1$, saddles -1), and their morphologies are shown in figure (1.3). Singularities of higher Poincaré index are possible [FG82], but are not generic (see next section).

It is easily seen that the gradient of phase $\nabla\chi$ in equation (1.2.3) has the form of a right-handed circulation (1.2.9) (its conjugate is left-handed), so phase singularities are circulations in the phase gradient field. Usually, functions cannot have critical points (stationary points, or zeros of gradient field) which are circulations, since gradient is curl free by Stokes' theorem, but phase, having values which are angles, can, and the phase singularities are circulations of phase gradient (hence the term *optical vortices* for their occurrence in optics [Sos98, SV01, VS99]). The two phase singularities on the globe (figure (1.2)) can therefore be interpreted as two counterrotating circulations (with vector flow lines along lines of latitude).

There are also critical points of phase χ where $\nabla\chi$ is a source or sink (local maximum, minimum of phase) or a saddle point (corresponding to a saddle points of phase). Phase singularities are also associated with sinks of $\nabla\rho$, since they are zeros of the nonnegative amplitude ρ. There are also local maxima and saddle points of ρ (sources and saddles of $\nabla\rho$); in general, the critical points of phase and amplitude are independent. Critical points of phase shall be considered further in the next chapter.

Related to vector singularities are singularities of *line fields*, that is fields of undirected flow lines ('headless vectors' [Mer79] or 'ridge systems' [Pen79]). Whereas vectors are only invariant with respect to a rotation by 2π, headless vectors is invariant with respect to a rotation by π. These fields are used to describe (in the continuum limit) nematic liquid
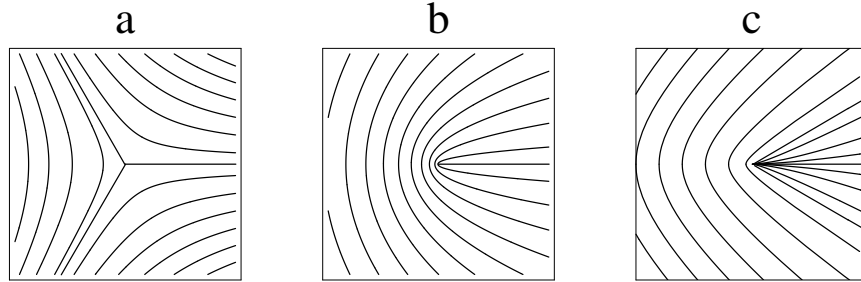
Figure 1.4: The three types of planar line field singularity with index $\pm 1/2$: (a) star; (b) lemon; (c) monstar. (Figure courtesy of Michael Berry.)

crystals (modelled by long ellipses) confined to a plane, and the direction (modulo π) in which the molecule points is called the *director*, and, as above, the angle of the director is described by the angle field $\theta = \theta(\mathbf{R})$, where θ is only defined mod π. Director fields can also have singularities, with index defined as for vectors by equation (1.2.7), only now, since θ is the same as $\theta + \pi$, the director rotates by an integer number of half-turns; the simplest director singularities have index $\pm 1/2$, and there are three distinct morphological types. They were discovered mathematically first by Darboux [Dar96] in the line fields of principal curvature of surfaces, and are depicted in figure (1.4). In this context, they are called *umbilic points*.

We shall follow Berry and Hannay [BH77] in calling these three singularities *lemon, star* and *monstar*. The lemon (L) is the most familiar singularity of line fields, being called a *disclination* (originally disinclination) in liquid crystal theory [Fra58] (the pattern is also called a 'loop' by [Pen79]). It has index $+1/2$, and has one straight line ending on the singularity. The star (S), by contrast, has index $-1/2$ and three lines ending on the singularity (it is called a 'triradius' by [Pen79]). The monstar (M) shares properties of both the lemon and star (hence the name (le)monstar), having index $+1/2$ but three locally straight lines meeting at the singular point.

These singularities shall play a very important part in the consideration of ellipse fields (usually the polarization ellipse in an electromagnetic vector field). In these fields (in two dimensions), there is an ellipse defined at each point, with varying orientation of major semiaxis (playing the role of the director), and eccentricity (size also can vary, but this is not important provided the ellipse does not vanish). However, when the eccentricity is 0, the ellipse is circular, and the ellipse orientation cannot be uniquely defined (for more

a   b   c
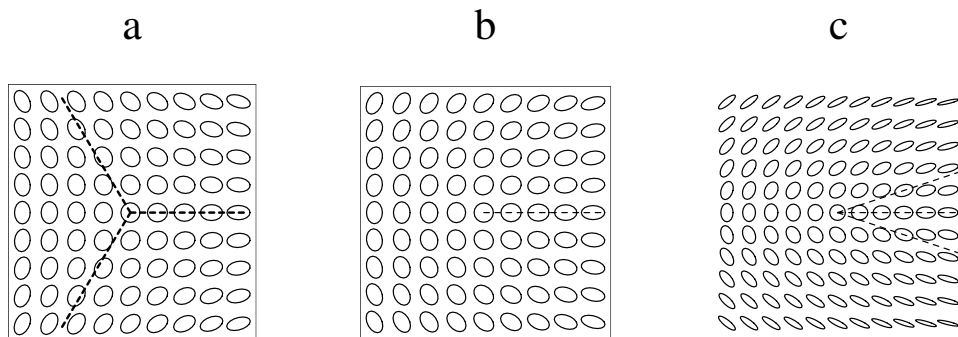
Figure 1.5: The three types of C point singularity in ellipse fields with index $\pm 1/2$: (a) star; (b) lemon; (c) monstar. The patterns are computed using (4.2.11) with $b = \pm 1$ for lemon/star, $b = 3$ for monstar.

details about ellipses and their geometry, see appendix A). Such points are called C points (circular points), and were first discussed in polarization fields by Nye [Nye83a], and the three types (lemon, star, monstar) are shown in ellipse fields in figure (1.5).

The singularities of lines of principal curvature considered by Darboux, are naturally described in terms of ellipses, as the Gauss curvature ellipse (with major axis in the direction of major curvature, the minor axis in the direction of minor curvature). For a real function $f(\mathbf{R})$, the lines of principal curvature are the eigenvectors of the hessian matrix $\partial_{ij}f$, with principal curvatures the eigenvalues, and umbilic points occur at degeneracies of this matrix.

1.3 Structural stability, codimension and catastrophe

One of the main reasons for studying singularities in waves is their ubiquity: they are structurally stable features of fields, and may be found in rather general wavefields (such as random waves), even when those fields are perturbed. Structural stability is often found in theories where symmetry does not play a role, and the best example (from which much of our terminology is drawn) is catastrophe theory [Arn86, PS78] and its application to geometrical and wave optics [Ber80, BU80, Nye99]. The concept of a general wavefield (to be made more precise in the following) is central to most of the work in this thesis, and is well exemplified by the random scalar wavefields of chapter 3 and vector wavefields of chapter 4. Chapter 4 of [PS78] is particularly relevant to the present discussion, complementing the descriptive account here with appropriate mathematical

terminology. An analysis of dislocations using catastrophe theory may be found in [Wri79].

In ray optics, a perfect point focus of a lens is not structurally stable, and if the lens is deformed slightly, the point explodes into a more complicated light pattern (a *caustic*), more or less dismissed until relatively recently as merely aberration. The unfolded forms (folds, cusps and the like) are themselves stable under further small perturbations, and are mathematically completely classified [PS78].

Zero points of complex fields in two dimensions (and lines in three) are similarly structurally stable; if a complex constant c is added to the field (1.2.3), the position of the zero moves to $(x, y) = -(\text{Re}\, c, \text{Im}\, c)$, but the zero does not vanish. This occurs similarly with the vector and line field singularities; under small changes, they almost always persist (excluding the possibility of annihilation between opposite charges/indices).

An important concept related to structural stability is *genericity*, which is another property possesses by our singularities: they occur naturally in fields, without further specific requirements. A related idea is that of *codimension*; the dimension of the singular locus is the dimension of the space (called the *control space* in catastrophe theory) minus the codimension of the singularity, which is two in the case of phase singularities (in the plane, $2 - 2 = 0$, a point singularity; in space, $3 - 2 = 1$, a line singularity). Morphological objects are often of a codimensional rather than a dimensional nature since the codimension conditions (such as $\xi = 0, \eta = 0$) are insensitive to the number of parameters specifying the control space. An example of a codimensional object from [PS78] is that of the border between two countries, which has codimension 1: on a two-dimensional map, the border is a line, but in three dimensions, one imagines it as a surface rising vertically from the ground (through which one passes even in an aeroplane).

The codimension of a phenomenon is computed from the number of independent homogeneous conditions required for it to occur. In the case of a phase singularity, $\psi = 0$ requires the real and imaginary parts ξ, η each to be zero; the zero contours of these functions are codimension 1 objects; in general circumstances, a real function in two dimensions vanishes along a line, in three dimensions on a surface. For both $\xi = 0, \eta = 0$ to be satisfied, the two zero contours must intersect, which happens along a line (mathematically, the contours correspond to manifolds, which are said to be *transverse*).

The zeros of real vector fields have a codimension equal to the dimension of the vectors; their conditions are that each of the components of the vector is zero. For vector fields, the dimension of whose vectors is equal to the dimension of the configuration space, the zeros

are always points. This shows why care must be taken in computing the codimension of a particular phenomenon: a zero of a real n-dimensional vector $\mathbf{v}$ is codimension n (corresponding to the n equations $v_i = 0$ for $i = 1, \ldots, n$), rather than codimension 1, which would appear to be the condition that the pythagorean length of the vector $|\mathbf{v}|$ be zero. In fact, this one condition is equivalent to the n components vanishing, since

$$|v|^2 = v_1^2 + \cdots + v_n^2, \tag{1.3.1}$$

which can only be zero if each of the n components v_i vanish. Codimension is computed from a smooth parameterisation of the object (the vector or scalar), in this case a cartesian decomposition with respect to a certain basis. The polar parameterisation of a vector is not smooth (when the length of the vector is zero, the polar angles are not defined), and is not appropriate for computing codimension (at least, in the vicinity of the problem point, as in the vector example). Since it is frequently these singularities of natural geometric parameterisations of objects that we are considering, we need to take particular care.

Another important example is in the case of polarization fields in two dimensions (as shall be discussed in detail in chapter 4). The field is described by a two-dimensional complex vector $\mathbf{E}$, which is smoothly parameterised by four real functions, the components of its real and imaginary parts with respect to some arbitrary, fixed basis. Geometrically, it represents an ellipse, which is traced out by the end of the vector $\mathrm{Re}\,\mathbf{E}\exp(-i\chi)$ as χ varies from 0 to 2π. The ellipse is described by four geometric variables (which can be found from the components of $\mathbf{E}$: its size, phase (position on the ellipse when $\chi = 0$), angle of the major semiaxis (with respect to a fixed direction), and the ellipse eccentricity ε (a detailed description of the geometry of ellipses may be found in appendix A). The ellipse field is singular (as in the previous section) when the ellipse is circular, since the major semiaxis is not defined, so the ellipse angle does not exist, occurring when $\varepsilon = 0$. Although this is only one condition, points of circularity (C points) are codimension two, as implied previously and to be shown later; moreover, the codimension of loci where the ellipse is linear is 1 (on L lines), although it is similarly a single condition on the eccentricity ($\varepsilon = 1$).

Generic features are often missed in concrete examples of physical phenomena, where, for mathematical or experimental convenience, the most symmetric configuration of apparatus is chosen (as in the example of a point focus above). However, generic features can most frequently be found when there are no manifest patterns imposed on the fields (the

dislocations in a sound field are expected to be generic whether the sound is background noise, or from a full orchestra playing a Beethoven symphony), and random waves, used to study the behaviour of singularities in chapter 3, are a good example of fields with the necessary lack of overall symmetry. The field structure is extremely complicated even if only a scalar field; for tensor fields, it is difficult even to visualise. However, the existence and morphology of the singularities provides a natural starting point for the description of these fields, and the singularities organise the structure of the parameter with respect to which they are singular; for instance, the phase structure of fields may be guessed from a knowledge merely of the location of the phase singularities. This is seen to be one of the major motivating factors for studying topological singularities in waves.

Although singularities are generic in random fields, locally, they themselves are highly symmetric, as seen by the local forms described in the previous section (saddles, sources, sinks or circulations for zeros in two-dimensional real vector fields, lemons, stars or monstars for C points). This local nature is important, and chapter 2 is devoted to elucidating this structure for phase singularities in two and three dimensions.

The connection between singularities and catastrophes in waves runs deeper than simply an analogy of mathematical description: dislocations play an important role in the structure of diffraction catastrophes [Ber91b, Ber92, Nye99, Wri77, BNW79], and are another good place to look for the generic behaviour of dislocations (although this is not investigated here). Berry [Ber94a, Ber98] also raises an interesting point to do with the nature of singularities in physics (optics, at least); in cases of waves where wavelength is small, but not vanishingly so, geometrical optics (where wavelength is nothing, and the waves are rays) is not sufficient, particularly at the caustics, where the geometrical intensity ought to be infinite, but is blurred out by interference, a feature of the wave nature of the light; the singularity is described by the two theories on either side of the physical limit being taken. He conjectures that a similar role might be played by dislocations, which are the singularities of wave optics; since they are zeros of the wave intensity, they could be places where photonic fluctuation may be detected (due to quantum optics, the next theory in the hierarchy describing light at different scales). Unfortunately, this very interesting question will not be investigated here.

1.4 Geometric phases and topological defects

Other than catastrophe theory, the theory of topological singularities in wave fields has obvious connections with two other theories in physics, those of topological defects in ordered media, and geometric phases.

The theory of topological defects is a particularly successful descriptive application of algebraic topology to problems in condensed matter (see [Mer79] for an excellent review). Normally, a condensed matter system comprises of a configuration (control) space, like two or three dimensional real space, which parameterises an order parameter, a mathematical object residing in a certain topological space (such as the complex plane). We have already seen this in the case of liquid crystals, which in the continuum limit are described by the orientations of ellipses in two dimensions, or ellipsoids in three approximating the shape of the liquid crystal molecule. These order parameter fields are geometrically analogous to the wave fields which are our concern here, although their mathematical nature (what equation the order parameter satisfies with respect to the control parameters), and physical nature (what the order parameter actually describes) may be quite different in the two cases.

Mathematically, the possible defects of the ordered medium are identified using the fundamental group of the order parameter as a topological space. The fundamental group is the group corresponding to closed paths (embeddings of the circle $\mathcal{S}_1$) in the order parameter space, identified by continuous transformation (homotopy), with composition given by the natural join of the two loops (via base-point homotopies), and is abelian (commutative). The fundamental group of the circle (taking phase, or polar angle of a real vector in the plane) is the group of (signed) integers $\mathbb{Z}$, the winding numbers of the circle corresponding to the topological charge of the phase singularity or Poincaré index of the vector field singularity. Note that it is important what the order parameter is that one chooses; if instead of the polar angle, the entire space $\mathbb{R}^2$ of planar vectors was chosen, there are no defects. Once again, the defects, as singularities, appear in the geometric parameters one chooses.

Examples of ordered media with a complex scalar order parameter are superfluid helium-4, or a singlet superconductor. The zeros of the superfluid helium order parameter are lines in space, called vortices, and their quantisation is well understood [TT90]. Topologically, dislocations as phase singularities are descriptively identical to superfluid

vortices, and many of the properties of phase singularities described here (particularly in chapters 2 and 5), such as the reconnection of crossing vortex filaments, applies to defects in appropriate media.

The topological description becomes important when the topology of the order parameter space is more complicated. For instance, the space of possible orientations of a three-dimensional (nonsymmetrical) object is the space of three-dimensional rotations SO(3), which is topologically identical to the three-dimensional projective plane, with fundamental group the two element group; this is a topological account of the fact that (tethered) rotations by 2π are not identified, but rotations by 4π are [Alt86, Fra97, Han98a]. This implies that defects in media with this order parameter (such as a field of orientations of symmetry-free molecules) can only have an index of $+1$, and if two are combined, they annihilate.

This shows a limitation of the topological defect theory when the order parameter is an object derived from the system (like orientation angles), at least for singularities in waves: the fundamental group description of the oriented ellipses in three dimensional polarization fields (as defined in chapter 4) suggests that two C lines, each of index $1/2$, could combine to form an index 1 object, suggested by the topology to be an L line. This does not seem to be the case, for reasons deeper than index alone. Topological reasoning alone is not sufficient to analyse singularities in waves.

Geometric phases have a lot in common with phase and polarization singularities [Ber91a, Ber91b, Nye91]. If the phase in a field is only defined up to a phase difference between neighbouring points, as in a quantum mechanical wavefunction [Dir31], then around a loop in control space, there is a net phase difference depending only on the geometry the loop and the field around (through) it [Ber84, SW89]. The control space may be mathematically rather complicated, although in simple cases, it may be two or three dimensional real space (the importance is in the phase connection between neighbouring points). Since, apart from the singularities, phase is well-defined everywhere in our wavefields, the only phase ambiguity between neighbouring points is at phase singularities (zeros), which topologically have codimension 2 from the genericity arguments of the previous section. Therefore, the quantised phase around a phase singularity is a special case of a topological (rather than geometric) phase.

It appears that Dirac [Dir31] was one of the first to appreciate the topological phase nature, when he suggested the quantisation of phase around nodal lines in quantum me-

chanical wavefunctions (and the resulting possibility of magnetic monopoles), although he missed the more general geometric phase possibilities. The Aharonov-Bohm effect [AB59] is another example of a topological phase, where the organisation of the vector potential into a circulation pattern around an infinitely thin, infinitely long solenoid giving rise to interference between two electron beams sent around each side of the solenoid appears to be a violation of locality (since the magnetic field is restricted to being within the solenoid). Alternatively, it is another example of the phase organising effect of a singularity (the magnetic vector potential is a circulation, the phase given by a line integral around it), and the presence of a phase singularity along the flux line was confirmed by [BCL$^+$80].

Another anticipation of the geometric phase was made by Pancharatnam [Pan56], who found the phase difference between two (two-dimensional) polarization states was dependent on the (closed) path of polarizations on the Poincaré sphere (see section A.4). The geometric connection of this Pancharatnam phase difference is the natural definition of propagation in three dimensional polarization fields [NH87, Nye91], and shall be shown, in fact, to be the expectation value of the local momentum operator in such fields (section 4.4 [BD01c]). It also provides an illustration of the difference between topological singularities and geometric phases, for polarization in planar paraxial fields: the Pancharatnam phase is calculated from the solid angle of the path on the Poincaré sphere corresponding to the sequence of polarizations the state passes through; the C point index counts the number of times the north pole is encircled by the path.

1.5 Waves, wavefields and wave equations

The fields with which we are concerned shall usually be solutions of a wave equation, and in this section we briefly review the relevant types of wave equation and solution, beginning with scalar waves. The wave equation usually imagined is the time-dependent (D'Alembert) wave equation

$$\nabla^2 \psi(\mathbf{r}, t) = 1/c^2 \ \partial_t^2 \psi(\mathbf{r}, t), \tag{1.5.1}$$

where $\mathbf{r}$ is position in two or three dimensions, and c is the wave speed. All of the waves we shall use are in free space, and no interactions, or boundary conditions of any kind are considered (so there are no evanescent waves [Ber94b]). We also ignore the technical

problems of whether the functional solutions of (1.5.1) and other wave equations are integrable (or square-integrable), and shall assume that Fourier transforms exist where necessary.

If the (complex) solution ψ to (1.5.1) is separable, the general solution of the time-dependent part is $\exp(-\mathrm{i}\omega t)$ (ignoring, for convenience, the possibility $\exp(\mathrm{i}\omega t)$) and the space-dependent part ψ_H satisfies the *Helmholtz equation*,

$$\nabla^2 \psi_\mathrm{H} + k^2 \psi_\mathrm{H} = 0, \tag{1.5.2}$$

where the real wavenumber k and angular frequency ω are of course related to the wave speed c by $c^2 = \omega^2/k^2$. Such solutions are called *monochromatic*: there is only one frequency component, and the pattern is periodic in time. Moreover, in monochromatic waves, zeros in the solution $\psi = \psi_\mathrm{H}\exp(-\mathrm{i}\omega t)$ of the D'Alembert equation (1.5.1) are exactly the zeros of the Helmholtz solution ψ_H, because the time dependent part $\exp(-\mathrm{i}\omega t)$ can never be zero.

The simplest, and most important, solution to the time-dependent wave equation (and also the Helmholtz equation, without t-dependence) is the *plane wave solution*

$$\psi = a\exp(\mathrm{i}(\mathbf{k}\cdot\mathbf{r} - \omega t)) \tag{1.5.3}$$

where a is some complex constant (the modulus $|a|$ being the plane wave amplitude, the argument being the phase). ω is the angular frequency, and $\mathbf{k}$ is the *wavevector*, which can be in any direction, provided its squared length $k^2 = \omega^2/c^2$. It represents a plane wave travelling in the $\mathbf{k}$-direction, in either two or three dimensional space. There are no singularities at all in the plane wave (1.5.3), and most authors who only identify plane wave solutions miss the singular wave structure that is present in superpositions of plane waves.

Fourier's theorem says that general solutions of the D'Alembert equation are, in fact, (usually infinite) superpositions of plane waves (1.5.3), labelled by their wavevectors $\mathbf{k}$, and the Fourier transform of ψ gives the complex wave amplitude a for the appropriate $\mathbf{k}$ labels. We shall usually write the Fourier decomposition as a sum of plane waves (labelled by the wavevector $\mathbf{k}$, as in (3.1.1), although it is in general an integral). Of course, general solutions of the Helmholtz equation (1.5.2) are also superpositions of time independent plane waves (of the form (1.5.3) with $\omega = 0$).

The wavefields in which phase singularities are to be found are general superpositions of plane waves. Their morphological nature is a feature of the particular plane waves added,

and a different choice of complex amplitudes $a_{\mathbf{k}}$ leads to very different fields, with totally different singularity structures. Structural stability, of course, implies that a small change does not remove the singularities, but possibly moves them. The nature of singularities in isotropic random superpositions of plane waves is the remit of chapter 3; the statistical properties of the singularities are found to be dependent only on the spectrum of the waves, that is, the distribution of the (real) square amplitudes with $k = |\mathbf{k}|$.

Other wave equations, such as the time-independent Schrödinger equation in quantum mechanics, also have Fourier components satisfying the Helmholtz equation; eigenfunctions in quantum billiards [Ber77] are an example of this (the boundary conditions in this case make the solution nontrivial).

Any plane section of a three-dimensional spatial wavefield is also obviously a wavefield, although its spectral decomposition will appear to be different from the true spatial spectrum of the wave. An important example of this is in the case of *paraxial waves*, for which the wavevectors of a three dimensional superposition are all in the z-direction (say), and, for simplicity, the overall wave is monochromatic. If $k \approx k_z$, and $\tilde{\psi} = \psi \exp(-\mathrm{i}k_z z)$ ($\tilde{\psi}$ in chapter 5 is called a *beam solution*), then

$$\nabla^2 \tilde{\psi} + (k^2 - k_z^2)\tilde{\psi} + 2\mathrm{i}k_z \partial_z \tilde{\psi} = 0, \tag{1.5.4}$$

and, if $|\partial_{zz}\tilde{\psi}| \ll 2|k_z \partial_z \tilde{\psi}|$, approximating k with k_z (the *paraxial approximation*), then $\tilde{\psi} = \psi_{\mathrm{P}}$ satisfies the *paraxial wave equation*

$$\nabla_\perp^2 \psi_{\mathrm{P}} + 2\mathrm{i}k\partial_z \psi_{\mathrm{P}} = 0, \tag{1.5.5}$$

where $\nabla_\perp^2 = \partial_{xx} + \partial_{yy}$ (the transverse laplacian). Most solutions of laser beams, because of their definite propagation direction, are considered to satisfy (1.5.5) [MW95]. The relation between paraxial and nonparaxial equations and their solutions is discussed in section 5.7; we observe here that the parabolic equation (1.5.5) is formally equivalent to the time-dependent Schrödinger equation in two spatial dimensions. Its Fourier components (in x, y) are two-dimensional plane waves, with transverse wavevectors $\mathbf{K} = (k_x, k_y)$ of variable length although the original three dimensional wave was monochromatic. Waves of this type are considered in chapter 3.

The fact that the wavefield ψ is complex is of crucial importance mathematically to us, since the phase singularities are essentially complex objects. How is the complexness explained physically? The answer is different, depending on the particular physical situation of the wave. If time-reversal symmetry is broken in the wave (such as in the tides, due

to the rotation of the earth [Ber01a], in quantum mechanics by a magnetic field [RB85], or in living systems [Win80]), the waves are naturally complex (even if the real part is the physical disturbance), and only the phase singularities are always zero (time-invariant). This is also the case for monochromatic waves; especially at the high frequencies of optics, only the time-invariant zeros can be easily measured. Sometimes the waves in the general superposition are all real (such as the case with billiard eigenfunctions; in the absence of a magnetic field, they are time-reversal symmetric and real), but their complex superpositions are still generic, and phase singularities may be found in them.

For other waves (such as nonmonochromatic, physically real solutions of (1.5.1)), more elaborate methods are necessary to construct a complex wave. The usual one used (and used here) is the so called *complex analytic signal* representation of the field, first introduced by Gabor [Gab46]. Such fields must have nontrivial time dependence, which is used in the construction. We wish to complexify the real field $\xi(t)$ to a complex field $\psi(t) = \xi(t) + \mathrm{i}\eta(t)$; clearly any function η, could be used, but is there a natural choice? The answer is yes: since ξ is real, its Fourier transform $\hat{\xi}$ must be symmetric in ω, and for every positive frequency component there is also a negative one. No information is lost if these are suppressed; the complex analytic signal ψ is the inverse Fourier transform of twice the positive frequency parts of $\hat{\xi}$, ie

$$\psi(t) = \frac{1}{2\pi} \int \mathrm{d}\omega (1 + \mathrm{sign}\,\omega)\hat{\xi}(\omega) \exp(-\mathrm{i}\omega t). \tag{1.5.6}$$

The imaginary part η is now readily identified as the Hilbert transform [Tit48] of ξ,

$$\eta(t) = \frac{1}{\pi} \fint \mathrm{d}t' \frac{\xi(t')}{t' - t}, \tag{1.5.7}$$

where $\fint$ represents the Cauchy principal value integral with pole at $t' = t$. If t is replaced by a complex variable, the function (1.5.6) is analytic (for analytic ξ), hence the term complex analytic signal.

The complex analytic signal obviously reduces to the earlier monochromatic case if ξ is a superposition of cosines; suppressing the negative frequency component gives a single δ-function, Fourier represented by an exponential $\exp(\mathrm{i}\omega t)$, and the Hilbert transform (1.5.7) of $\cos\omega t$ is $\sin\omega t$. The complex analytic signal is also useful in quantum optics, where the complex ψ is second quantised to the field annihilation operator $\widehat{\psi}$, and for narrow-band signals defines an envelope (the amplitude ρ (1.2.1)) with minimal fluctuations [Man67] (and probably, minimal number of phase singularities).

Recently, a two-dimensional spatial generalisation of the Hilbert transform (1.5.7) has been proposed [LBO01, Lar01], where the sign function is replaced by the function (1.2.3), with a phase singularity at the origin. This provides a way to complexify spatial fields independently of their time components, but this has yet to be applied to fields containing phase singularities.

All of the above comments generalise directly to vector or tensor fields; the components, in three dimensions, of vector or tensor waves satisfy an appropriate wave equation. Transverseness (that is, the vector or tensor field is divergenceless, see section 6.4 for a general discussion), equivalent to requiring that the wave disturbance of each plane wave component is orthogonal to its wavevector, is always imposed on the tensor or vector fields we study here. Relativistically, this corresponds to the fact that all fields we study correspond to massless particles [Wig39, FMW99], since the vector and tensor wavefields satisfy the time-dependent wave equation (1.5.1) [BW48]. This implies that paraxial vector waves are confined to the (x, y) plane, as is assumed in chapter 4.

1.6   A brief history of phase singularities

It is hardly surprising that phase singularities in waves were found shortly after wave interference began to be studied intensively (following Young's observation that light interferes, so is a wave [Par97]). Since then they have been rediscovered several times in different physical contexts, before being cast in a general wave framework by Nye and Berry in 1974 [NB74]. This has been reviewed before, (see for example [Ber81, Ber01a]). More recent developments, where relevant to this work, are cited in the appropriate place, and also in [Nye99]. The place of topological singularities in waves in the general framework of Bristol geometric physics has been discussed by Berry [Ber91a].

It appears that the first phase singularities in waves to be discovered physically are the so-called amphidromic points in the tides. Although the tidal wave equations, governing the height of water at a given position and time, due to Newton and Laplace, were well-known at the beginning of the nineteenth century [Car99], applying the theory to the real tides, for example in the North Sea, is an analytically impossible problem, due to the rather complicated coastal boundary conditions. Whewell [Whe33] wished to understand the tides empirically by computing a "map of cotidal lines", showing the various equiphase lines of the complex wave amplitude (complex due to the broken time-reversal symmetry

of the rotation of the earth, although, of course, periodic).

He found [Whe36] that, for the pattern of lines in the North Sea,

> we may best combine all the facts into a consistent scheme, by dividing this ocean [the North Sea] into two *rotatory* systems of tide waves, one occupying [the region] from Norfolk and Holland to Norway; and the other the space .. between the Netherlands and England... The cotidal lines may be supposed to revolve around a point ... where there is no tide; for it is clear that at a point where all the cotidal lines meet, it is high water equally at all hours, that is, the tide vanishes... [The southern one] resembles a watch or clock, which is kept in continual motion by a sustaining force applied at intervals.[2]

What he is describing is clearly a phase singularity, and he identifies many of the features discussed in section 1.2: all the equiphase lines meet at the singularity, it is a (perpetual) node of the wave, the pattern rotates around the singularity. The phase singularity between Norfolk and the Netherlands, adapted from Whewell's original cotidal line map, is given in figure (1.6). Initially Whewell's interpretation of the data was contested by Airy, but soon became widely accepted, and identification of the amphidromic points is a major part of the theory of tides [Car99].

The next appearance of phase singularities in waves seems to have been in Dirac's landmark paper [Dir31], in which he not only observes that a (complex three-dimensional) quantum wavefunction must have nodal lines along which the phase is singular, but uses Stokes' theorem to show that the string must end on a magnetic monopole-like object. As with the Aharonov-Bohm effect, this is done by realising that the electric vector potential $\mathbf{A} = \hbar c/e \nabla \chi$ in the Schrödinger equation for a charged particle with charge e. If the line of phase singularity ('Dirac string') ends (only theoretically possible because the wavefunction phase can only be defined between neighbouring points), around a vanishingly small path C enclosing the phase singularity of nonzero strength s,

$$
\begin{aligned}
2\pi s &= \oint_C \mathrm{d}\mathbf{r} \cdot \nabla \chi = \frac{e}{\hbar c} \oint_C \mathrm{d}\mathbf{r} \cdot \mathbf{A} \\
&= \frac{e}{\hbar c} \int_{\mathcal{S}} \mathrm{d}^2 u \, \nabla \wedge \mathbf{A} = \frac{e}{\hbar c} \int_{\mathcal{S}} \mathrm{d}^2 u \, \mathbf{H} \\
&= 4\pi \mu \frac{e}{\hbar c}
\end{aligned}
\tag{1.6.1}
$$

---

[2]Quotation from [Whe36], pages 298-299.
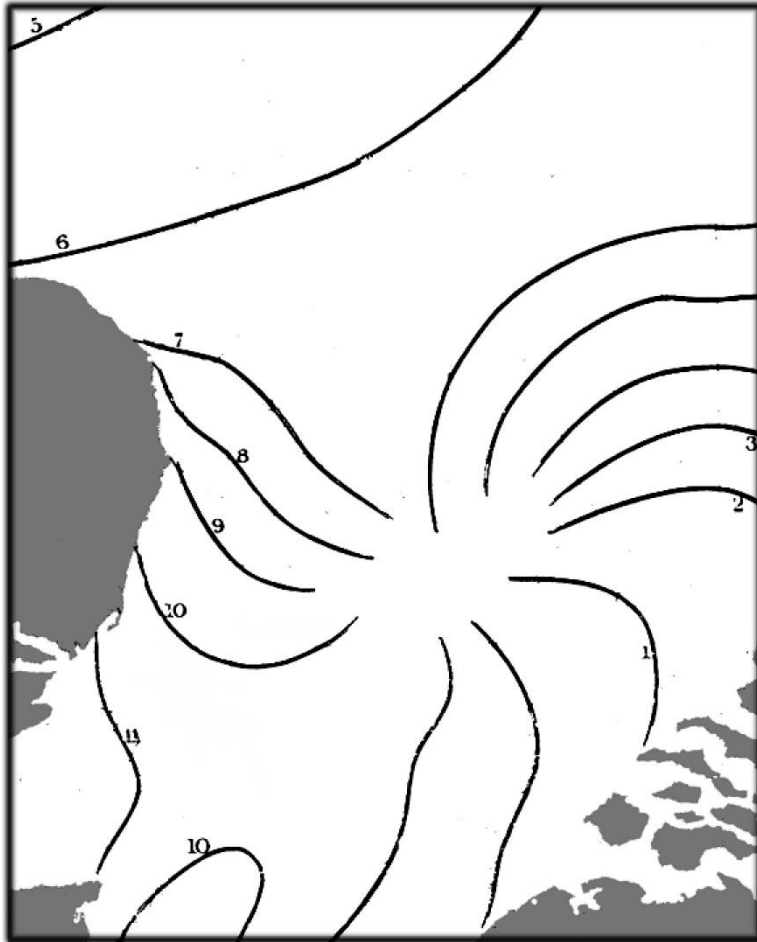
Figure 1.6: Showing the amphidromic point (phase singularity) in the tide in the North Sea between Norfolk and Holland, adapted from [Whe36]. The numbers on the cotidal lines indicate the phase, that is, the hour at which that line is the high tide. (Figure courtesy of Michael Berry.)

where Stokes' theorem has been applied in the third line over some surface $\mathcal{S}$ with boundary C enclosing the string end point, and $\mathbf{H}$ is the magnetic field, with a nonzero flux μ through $\mathcal{S}$. The product of electric and magnetic charge is therefore quantised topologically, in units of $\hbar c/2$.

Dirac himself realised that there is a problem with the quantum phase singularities under a (noncontinuous) gauge transformation: the phase singularity can be moved, but the node of amplitude cannot (since it is observable). This problem was later resolved by Wu and Yang [WY75], but only by generalising the notion of magnetic vector potential using fibre bundles, and removing the phase singularity altogether. They also investigate what monopole-like objects there are in gauge fields with symmetries other than U(1) (phase), concluding that the possible monopole charges are those of the fundamental group of the topological gauge group (in a similar way to defects of the order parameter, as discussed in section 1.4). In particular, there are no monopoles in SU(2)-symmetric fields; this is a version of the fact, discussed in chapter 6, that there are no topological singularities in wavefunctions of particles of spin $1/2$ (such as neutrinos).

In the 1970s, there was renewed interest in phase singularities in waves. In quantum mechanical wavefunctions, they were studied by Hirschfelder and collaborators [HCP75, HGB74, HT76a, HT76b], in which the singularities are identified with circulations of the current (discussed here in section 2.2). Riess [Rie70a, Rie70b, Rie87] has investigated the nodal structure of multiparticle solutions of the Schrödinger equation. Also, Winfree [Win80] worked on phase singularities in biological systems where cyclic rhythms are important. The starting point for the work described here was that of Nye and Berry [NB74], who noticed the analogy between nodal points and lines and crystal dislocations.[3]

1.7 A brief history of polarization singularities

Although polarization singularities in wavefields are not as ubiquitous as phase singularities (fields with a vector nature are more difficult to come by, and it is easier to study a scalar component of a polarization field than the whole field), the first examples of polarization singularities were discovered before phase singularities.

---

[3]Michael Berry informs me that this paper was initially rejected; one referee claimed that the ideas were too simple. They are indeed simple, but we hope that this thesis shows they sometimes have surprising subtleties too.
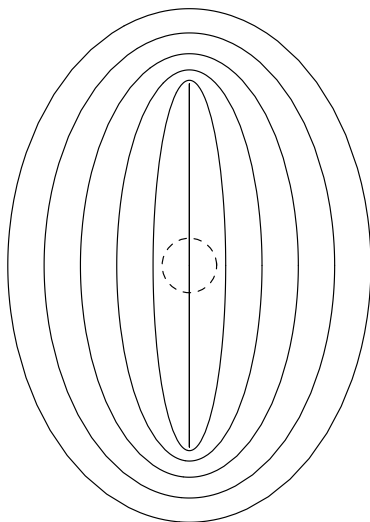
Figure 1.7: The polarization pattern around the sun (or antisun), drawn as a dotted circle. The polarization at a point is tangent to the line at that point; there are two lemon-like singularities, one above and one below the sun/antisun.

   Shortly after Malus discovered the polarizable nature of natural light [Bro98], Arago discovered (in addition to the laws concerning interference of polarized light) that sunlight in the sky is naturally polarized, but there is a point in the polarization pattern of sunlight in the sky of no polarization; Babinet and Brewster later found three more [Bre47, Lee98, GHMRW01]. The nonzero polarization of skylight is due to scattering, and is linear in the direction perpendicular to the plane of the incident and scattered ray, and the polarization in a given direction is the sum of all contributions from the different scattered rays finishing in that direction, and are only partially polarized (states are described by a point within the Poincaré sphere (appendix A)); the polarization state is therefore restricted to the equatorial disk of the Poincaré sphere. The four neutral points are all of lemon type, and all lie on the great circle in the sky including the zenith and the sun. The four points are arranged near the position of the sun and the antisolar point (or antisun, ie the point in the sky antipodal to the sun), and the pattern of linear polarizations around these two points is shown schematically in figure (1.7), and experimentally measured polarization maps may be found in [HGP98].

   These singularities may be easily understood from a perturbation argument. The strongest contribution is from rays that are only scattered once, the polarization pattern

from this clearly being a series of concentric circles around the sun/antisun (rather than confocal ellipses, as in the figure), with a point of no polarization in the (anti)solar direction. As we have seen, a singularity in a pattern of lines generically has index $\pm 1/2$, rather than 1 (as in this case), and this unstable configuration is perturbed by secondary scatter. Although the secondary scatter is weaker, it is sufficient to perturb this pattern, and the direction of polarization from rays that are scattered in the atmosphere twice, in the solar direction, is vertical, since more scattering takes place in the directions to the left or right of the (anti)solar point rather than above or below it. More detailed calculations of this phenomenon may be found in [Cha50, vdH49].

Although Rayleigh [Ray71] described the phenomenon of polarization in the sky, he neglected to mention the polarization neutral points in the sky, and these appear to have been more or less forgotten by non-atmospheric physicists for a while. In the 1970s, however, a group of astronomers (including Hannay) discovered polarization neutral points in patterns of radiation from stars [SHH77], where only the star and lemon were identified, with the monstar coming later [BH77]. Polarization singularities of the types described here were found to occur generally in paraxial polarization fields by [Nye83a], and in three dimensional fields by [NH87].

The tides also provide a two dimensional ellipse field - that of the tidal current, whose vector traces out an ellipse in time [Car99]. Berry [Ber01a] has recently conjectured the positions of tidal polarization singularities in the North Sea.

1.8   Outline of thesis

The layout of the thesis is as follows.

The second chapter is concerned with questions of dislocation geometry and morphology, in both two and three dimensions. The topology of dislocations on contour lines (including the so-called sign rule) is described, and the interrelation between phase singularities and phase critical points, put on an equal footing as zeros of the field current, is discussed. The local structure and motion of dislocations are also investigated. In three dimensions, the reconnection of dislocation lines is described, and general expressions are derived for dislocation curvature and torsion, as well the generalisation of the geometric structures of two-dimensional dislocation points. The chapter concludes with a section on the geometry of the core structure twisting around the dislocation line, and several

alternative measures of this are constructed.

The third chapter is wholly statistical, and the model used is an ensemble of isotropic gaussian random waves, made up from superpositions of plane waves isotropically distributed in direction, but with random phases. As well as calculating statistical averages and probability densities of the geometric quantities defined in chapter 2, the charge and number correlation functions of dislocation points in the plane (or plane section) are calculated and discussed in detail.

The subject of the fourth chapter is polarization singularities in vector waves, whose structure in two and three dimensions is described, and its relevance to electromagnetic singularities discussed. Calculations of statistical densities of polarization singularities in gaussian random vector waves are made.

In the fifth chapter, the topology of closed dislocation loops is used to construct a family of monochromatic waves, in which the dislocation lines are knotted and linked. Explicit constructions of the trefoil knot and Hopf link are derived in experimentally realisable Bessel and Laguerre-Gauss wave beams. The creation and dissolution of knots is investigated, using the reconnection mechanism described in the second chapter.

The final chapter is a discussion on how the polarization singularities of vector waves generalise to fields of spinors, using the Majorana sphere, and the resulting structures are similar, but more complicated geometrically in the three-dimensional case. The particular case of fields of gravitational waves, which are described by a traceless complex symmetric matrix at each point, are considered briefly.

There is one appendix, which describes some well-known (and some less well-known) aspects of ellipse geometry; ellipses play a central role in almost all of the material.

There is no overall conclusions section; the outline above is an adequate summary of the work presented in this thesis, and each chapter concludes with a summarising discussion, including suggestions for future investigation.

The layout, as described above, may appear somewhat haphazard; phase singularities are returned to in chapter 5 just after polarization singularities are introduced, which investigated further in chapter 6. This is (roughly) the order in which the work reported here was done, and there are conceptual threads relating the different chapters, although each chapter is self contained. Most of the work has already been published, in [BD00, BD01c, Den01a, Den01b, BD01a, BD01b], and the reader is referred to these for further details.

1.9 Notation conventions

We hope that the notation used in this thesis is sufficiently unremarkable that confusion will not ensue. Due to the range of physics and mathematics discussed, it is impossible to avoid certain notational clashes (such as ω, representing an angular frequency and also vorticity, or ε representing variously ellipse eccentricity, random wave amplitude, the antisymmetric symbol or a small perturbation parameter). The appropriate meaning is (hopefully) clear from the context.

Use is made of the three standard coordinate systems in three-dimensional space: cartesian coordinates (x, y, z), cylindrical coordinates (R, ϕ, z), and spherical coordinates (R, θ, ϕ). ϕ is used to represent the azimuthal angle, and the polar angle in two dimensions. Vectors in the plane are usually denoted by uppercase letters (such as the position vector $\mathbf{R}$), whereas vectors in three dimensions are lowercase (as with the three dimensional position vector $\mathbf{r}$). It is not always possible to adhere to these conventions; unit basis vectors (such as $\mathbf{e}_x, \mathbf{e}_y$) are always lowercase, and the electric field vector $\mathbf{E}$ is always uppercase. The modulus of a (real) vector is usually given the same symbol, but in italic rather than bold (eg $|\mathbf{p}| \equiv p$). The symbol $\wedge$ is used for the vector product; the symbol $\times$ is used to denote a multiplication broken over two lines (as in the third line of equation (2.8.7)).

Partial derivatives of scalars are usually written as suffixes (eg $\partial \psi / \partial x \equiv \psi_x$), and partial derivatives of vector components as suffixes preceded by a comma (eg $\partial v_i / \partial x \equiv v_{i,x}$). Use is made of the summation convention (ie summing over repeated indices), as usually indicated in the text.

When integrals do not have limits, it is understood that they run from $-\infty$ to ∞, and never represent an indefinite integral. $\fint$ represents a Cauchy principal value integral, and the position of the pole will always be made specific.

Gauge transformations usually refer to a global phase transformation (such as $\psi \rightarrow \psi \exp(i\phi_0)$, with ϕ_0 constant); the context is clear when a different type of gauge transformation is appropriate. The spinor notation used in chapter 6 is explained in section 6.1.

Chapter 2

Phase singularity morphology and geometry

> 'We expect to find a hole in the theory here ... a naked singularity would
> be *very* messy. The mathematics is inconsistent - like dividing zero by zero.'
>
> Larry Niven, *Singularities Make Me Nervous*, in *Convergent Series*, Macdonald Futura,
> 1980

In this chapter we investigate the detailed mathematical structure of phase singularities, with emphasis on the physical features of generic dislocations in scalar waves. The first four sections deal with singular points in two dimensional scalar fields, the others the geometry and topology of singular lines in three dimensions. As well as setting the scene mathematically, this chapter is related to the rest of the thesis in the following ways: in chapter 3 averages of quantities calculated here are computed in isotropic random waves, and the geometry and topology of dislocation lines explained in this chapter are used in the knotting constructions of chapter 5. The other chapters are concerned with the vector theory (chapter 4) and tensor theory (chapter 6) of singularities, which are best understood with the help of scalar singularities. Examples of waves with dislocations exhibiting the properties discussed are given where appropriate. Much of the earlier material in this chapter is well-known dislocation theory, and some of the original results have been published in [BD00, BD01a, BD01b, Den01a, Den01b].

2.1   Dislocation strength and level crossing topology in two dimensions

As we have seen, phase singularities in complex scalar wavefields ψ occur at the zeros $\psi = 0$, that is, on the crossings of the zero contours (level sets) of the real and imaginary parts ξ, η of ψ (equations (1.2.1), (1.2.2)). We have also seen that in order to be a topological singularity, the phase (argument) χ around a zero must change by a signed nonzero integer multiple of 2π, which is called the topological charge (dislocation strength) s, positive if the phase increases in an anticlockwise sense with respect to a circuit round the dislocation, negative if clockwise (equation (1.2.4)). Also, although all phase lines (contours of χ) meet at the singularity, all essential local information can be extracted from the contours of the real and imaginary parts (and their derivatives).

Although each phase line (mod 2π) ends on a dislocation, each of the real and imaginary zero contours (which follow the phases 0 and $\pi/2$, mod π) are unaffected by the precise position of the dislocation on them (this being dependent on the other contour), since the phase is defined on such a contour only up to a sign (ie mod π), and, as the singularity is crossed, the phase on the contour changes by π and the direction of the gradients $\nabla\xi, \nabla\eta$ switch direction. Phase contours (wavefronts) shall sometimes be used modulo π, and sometimes modulo 2π, depending on the context, usually mod π. The phase contours mod π do not end - in two dimensions, they are either extend to infinity or are closed (generically nonselfintersecting) loops; for a given phase χ, they are the zeros of the real function

$$u_\chi = \operatorname{Re}\psi \exp(-\mathrm{i}\chi). \tag{2.1.1}$$

ξ, η are two particular phase contours with phases 0, $\pi/2$ respectively, and in fact,

$$u_\chi = \xi \cos\chi + \eta \sin\chi. \tag{2.1.2}$$

This clarifies the requirement that all phase contours cross at a dislocation - since, by equation (2.1.2) only two distinct contours (for convenience, $\xi = 0, \eta = 0$) are required to cross (the singularity has codimension 2). At a singularity with a absolute value of strength greater than 1, the phase contours intersect themselves as well as each other, and the codimension and morphology of high strength dislocations is discussed in section 2.3. For the rest of this section, we shall only consider singularities of topological charge ± 1.

Changing coordinates so the dislocation is at the origin, ψ can be expanded

$$
\begin{aligned}
\psi &= \xi_x x + \xi_y y + \mathrm{i}(\eta_x x + \eta_y y) + \dots \\
&= \nabla \xi \cdot \mathbf{R} + \mathrm{i} \nabla \eta \cdot \mathbf{R} + \dots,
\end{aligned}
\tag{2.1.3}
$$

where derivatives ξ_x, etc are taken at 0. The neighbourhood of the origin in the (ξ, η) plane is mapped to the neighbourhood of the singularity in the (x, y) plane by the jacobian matrix $\mathbf{M}$,

$$
\mathbf{M} = \begin{pmatrix} \xi_x & \xi_y \\ \eta_x & \eta_y \end{pmatrix}.
\tag{2.1.4}
$$

The sign of $\det \mathbf{M}$ gives the sign of the topological charge (that is, the dislocation strength, since we assume the singularity is generic); a negative determinant means that $\mathbf{M}$ reverses handedness, so when a circuit is anticlockwise in (ξ, η) space, its image in (x, y) space is clockwise. For a generic singularity, therefore,

$$
\begin{aligned}
s &= \operatorname{sign} \det \mathbf{M} \\
&= \operatorname{sign}(\xi_x \eta_y - \xi_y \eta_x) \\
&= \operatorname{sign} \operatorname{Im} \nabla \psi^* \wedge \nabla \psi,
\end{aligned}
\tag{2.1.5}
$$

where the final equality expresses s is a form invariant of global choice of phase (gauge-invariant) [Ber98]; the second equality is the form of s one gets from equation (1.2.4) when χ is written in terms of ξ, η. The local matrix $\mathbf{M}$ is investigated further in section 2.3.

The number of (generic, strength ± 1) phase singularities $D_{\mathcal{A}}$ in an area $\mathcal{A}$ of the plane can be found by generalising the well-known expression for number of zeros of a one-dimensional function (see, for example, [Gri87] appendix A), and is [Ber78, BD00]

$$
D_{\mathcal{A}} = \int_{\mathcal{A}} \mathrm{d}^2 \mathbf{R} \, \delta(\xi) \delta(\eta) |\xi_x \eta_y - \xi_y \eta_x|
\tag{2.1.6}
$$

(the δ-functions pick out the zero contours, and the modulus of the jacobian is the correct factor in transforming from the origin of the (ξ, η) plane to the (x, y) plane, counting $+1$ for each singularity). If the dislocation has a strength of modulus higher than 1, the first derivatives $\nabla \xi, \nabla \eta$ are zero and the jacobian disappears. To find the total dislocation strength $s_{\mathcal{A}}$ in area $\mathcal{A}$, each singularity is weighted by its topological charge $s = \pm 1$, which is the sign of the jacobian (2.1.5), giving

$$
s_{\mathcal{A}} = \int_{\mathcal{A}} \mathrm{d}^2 \mathbf{R} \, \delta(\xi) \delta(\eta) (\xi_x \eta_y - \xi_y \eta_x).
\tag{2.1.7}
$$

If $\mathcal{A}$ is simply-connected with boundary C, equation (1.2.4) can be derived from equation (2.1.7) by Stokes' theorem.

The statement that the total topological charge of singularities in an area is given by the net change in argument around the area is reminiscent of the argument principle in complex analysis [Bea79] : for a well-behaved meromorphic complex function $w(z)$ with variable z complex, the difference between the number of zeros and poles in an area $\mathcal{A}$ is equal to the change in argument around the boundary of $\mathcal{A}$. This illustrates the contrast between ψ as a real analytic map $\mathbb{R}^2 \longrightarrow \mathbb{C}$ and w, a complex map $\mathbb{C} \longrightarrow \mathbb{C}$, analytic except at poles; for ψ to be analytic requires its real and imaginary parts ξ, η to be (independently) real analytic, and around a zero, are locally expanded as in equation (2.1.3). On the other hand, if w has a zero at the origin, it is expanded

$$w(z) = w'z + \ldots \tag{2.1.8}$$

(w' is the derivative with respect to z at 0), implying that the analogue of the local matrix $\mathbf{M}$ of (2.1.4) is the single complex number w'. The argument around the zero always increases anticlockwise, so zeros of w are always of positive strength, and the modulus of the function grows uniformly with distance from the zero (scaled by $|w'|$). Expanding about a (simple) pole at the origin,

$$w(z) = w'\frac{1}{z} + \ldots \tag{2.1.9}$$

and since $1/z = z^*/|z|^2$, the argument increases anticlockwise about the pole, analogous to a negative charged phase singularity. $\mathbf{M}$ has two additional degrees of freedom to w', related to the local dislocation structure (see section 2.3), which is richer than the structure of zeros and poles of meromorphic complex functions.

The two sets of zero contours $\xi = 0, \eta = 0$ partition the plane into four parts (the images of the four quadrants of the (ξ, η) plane, with phases in the four quadrants of χ). All four meet at dislocations, and an illustration of the ξ, η contours with dislocations at the crossings is given in figure (2.1). The topological fact that has come to be called the sign principle, first stated explicitly and explored thoroughly by Freund and coworkers (see, for example, [FS94, Fre97, Fre98b], and references therein) follows immediately, and can be stated in the following way: dislocation points adjacent on a zero contour of ξ, η (or more generally, any phase contour after appropriate gauge transformation) have opposite sign. The sign rule holds because the signs of the gradient of η along a ξ contour at dislocations adjacent on $\xi = 0$ are opposite.
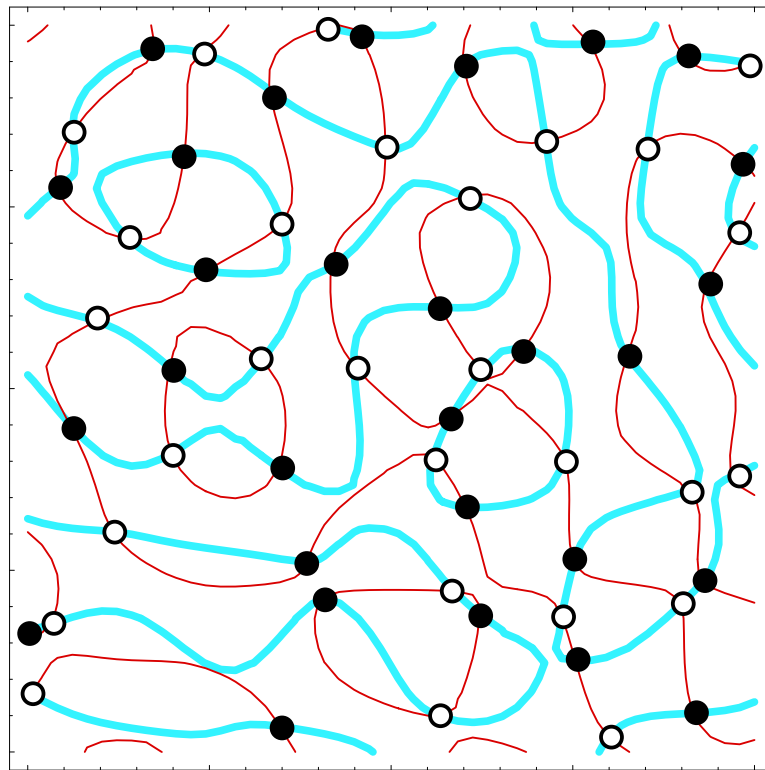
Figure 2.1: Illustrating the sign principle, in a random wavefield: the $\xi = 0$ contours are the thick lines, $\eta = 0$ the thin lines. Positive strength dislocations are the filled circles, negative ones are empty circles.

A given singularity is usually adjacent, on different phase contours, to different singularities, although the sign rule holds for any contour. This implies an overall anticorrelation in sign [SF94b, SF94a], discussed for random waves in the next chapter. If a phase contour is closed, there must be an even number of phase singularities, alternating in sign, and the overall topological charge on that contour is zero, giving an overall topological neutrality condition: topological charge cannot accumulate on a closed contour loop (this is clear from figure (2.1)).

The sign rule also illustrates a general and important fact: as external parameters vary (such as time, or the z axis for a plane section of a three dimensional wave pattern), singularities are created/annihilated in pairs of opposite strength as a ξ contour crosses an η contour (this is just happening at the lower left-hand corner of the figure, and is about to happen to the right of this, along the same ξ contour). This is the first singularity conservation law of several we shall see at various stages: in reactions, the total topological charge is conserved. This was, of course, implicit from the beginning, since provided no singularity crosses the loop C in (1.2.4), the topological charge s enclosed does not change under continuous variation of parameters.

An example of two adjacent dislocations with opposite signs in an exact solution of the Helmholtz equation is the modification of the edge dislocation of equation (1.2.5), discussed by [NB74, Nye99],

$$\psi = (K(x^2 - ax) + \mathrm{i}y)\exp(\mathrm{i}Ky), \qquad (2.1.10)$$

with zeros at $(0, 0)$ and $(a, 0)$, the second moving as the real parameter a is varied. Simple modifications of this wave (replacing y by $k(y^2 - by)$ in the left hand term) produces exact solutions of the Helmholtz equations with dislocations that are created/annihilated, as described in [NB74].

2.2   Current topology and phase critical points

As discussed in section 1.2, the phase χ not only has singularities, but also critical points where its gradient $\nabla\chi$ vanishes, which can be saddle points or extrema (maxima/minima). The (Poincaré) index of such a critical point (as a singularity of $\nabla\chi$) is the average of signs of the eigenvalues of the hessian matrix $\chi_{\alpha\beta}$ ($\alpha, \beta = x, y$), and maxima are distinguished from minima since both eigenvalues of the hessian are negative for a maximum, positive for

a minimum. Only critical points of index ± 1 are considered, since only these are generic; the hessian is always considered to be nonsingular. The intensity ρ^2 can also have critical points, investigated by [WH82, Fre96], but these are not considered here.

The current associated with ψ is the (two-dimensional) *current* $\mathbf{J}$ (properly *current density*), defined

$$\mathbf{J} = \mathrm{Im}\,\psi^*\nabla\psi = \rho^2\nabla\chi = \xi\nabla\eta - \eta\nabla\xi, \tag{2.2.1}$$

which is a real vector field, analogous to local momentum, satisfying the continuity equation (see, for example, [LL77]), and is invariant under global phase change. It is the probability current density if ψ is a quantum mechanical wavefunction (the local expectation value of momentum), and the Poynting vector in scalar theories of light (see, for instance [APB99]).

The *current vorticity* ω is defined

$$\omega = \frac{1}{2}\nabla \wedge \mathbf{J} = \frac{1}{2}\,\mathrm{Im}(\nabla\psi^* \wedge \psi) = \nabla\xi \wedge \nabla\eta. \tag{2.2.2}$$

It is properly a 2-form, which can be thought of, in two dimensions, as a real number, or the z-component of a vector perpendicular to the (x, y) plane (ie $\omega = \omega\cdot\mathbf{e}_z$)). It is invariant up to a global phase change, and, by the last equality, is equal to $\det\mathbf{M}$ of equation (2.1.4). Thus $\mathrm{sign}\,\omega$ is the topological charge of the singularity, and $\mathbf{J}$ circulates around the zero in the sense of its strength (as does $\nabla\chi$). Since the current plays an important role in light, it is more due to circulation of current than phase gradient that dislocations are called optical vortices. By equation (2.2.1), the current vanishes on a vortex, and does not have any singularities (it is zero whenever its direction is undefined).

This natural association of a nonsingular real vector field with the scalar field ψ has therefore connected two topological features: the phase singularities are current vortices. Taking such a point at the origin and expanding $\mathbf{J}$,

$$\begin{aligned}
\mathbf{J} &\approx (\nabla\xi \cdot \mathbf{R})\nabla\eta - (\nabla\eta \cdot \mathbf{R})\nabla\xi \\
&= \omega_0 \wedge \mathbf{R}, \tag{2.2.3}
\end{aligned}$$

so the current not only circulates in the direction of the topological charge, but the flow lines are, in fact, locally perfectly circular, however anisotropic the dislocation structure is ([Ber01a], fig (2) shows a particularly striking example of this). Mathematically, the form for $\mathbf{J}$ given by equation (2.2.1) is not completely general; as shown here, the circulations cannot be elliptical or spiral.

From equation (2.2.1), the other zeros of current are at the critical points of phase (where $\nabla\chi = 0$), where the fraction

$$\frac{\xi}{\eta} = \frac{\xi_x}{\eta_x} = \frac{\xi_y}{\eta_y} \qquad (2.2.4)$$

is generically nonzero and finite.

The conservation of index I_P (equation (1.2.7)) places additional restrictions on the allowed topological reactions between phase singularities: for two singularities (each with Poincaré index $+1$) to create/annihilate, not only must they be oppositely charged, but two saddles (with index -1) must also be present, so, where $D^\pm$ denotes a dislocation of charge ± 1, S a saddle, and $\leftrightharpoons$ the relation between reaction input and output,

$$D^+ + D^- + 2S \leftrightharpoons 0; \qquad (2.2.5)$$

the total topological charge and Poincaré index on each side being 0. This reaction was studied in detail for waves by [NHH88], who showed that, if the wave near the dislocation satisfies Laplace's equation, the four singularities must lie on a rectangle with dislocations on opposite corners in the creation/annihilation limit. An alternative reaction, mentioned at the end of appendix A of [NHH88] and investigated more thoroughly by Freund [Fre95, FK01] is

$$D^+ + D^- + S \leftrightharpoons M^+ + M^- + S \qquad (2.2.6)$$

(where M^+ denotes a phase maximum, M^- a phase minimum); the total index is $+1$ on each side, and it seems that this reaction can only take place when a 'reentrant' saddle is present [Fre95]. The conservation of index also holds for intensity critical point fields, and dislocation creation/annihilation usually involves two intensity saddles, analogous to the reaction (2.2.5).

Defining the matrix $\mathbf{M_J} \equiv \partial_\alpha J_\beta$, $(\alpha, \beta = x, y)$, the number of current zeros $Z_\mathcal{A}$ in area $\mathcal{A}$ is

$$Z_\mathcal{A} = \int_\mathcal{A} \mathrm{d}^2\mathbf{R}\,\delta^2(\mathbf{J})|\det \mathbf{M_J}| \qquad (2.2.7)$$

(replacing J_x, J_y for ξ, η in equation (2.1.6)). Since $\mathrm{sign}\det \mathbf{M_J}$ is the index I_P of the critical point, the total index $I_\mathcal{A}$ in $\mathcal{A}$ is (as with equation (2.1.7))

$$I_\mathcal{A} = \int_\mathcal{A} \mathrm{d}^2\mathbf{R}\,\delta^2(\mathbf{J})\det \mathbf{M_J}. \qquad (2.2.8)$$

By the difference between two squares,

$$\det \mathbf{M_J} = \det \mathbf{M_J}_{\mathrm{sym}} + \det \mathbf{M_J}_{\mathrm{asym}}, \qquad (2.2.9)$$

where $\mathbf{M_{J}}_{\text{sym}}, \mathbf{M_{J}}_{\text{asym}}$ are the symmetric and antisymmetric parts respectively of $\mathbf{M_J}$, with determinants

$$\det \mathbf{M_{J}}_{\text{sym}} = \frac{1}{2}(\nabla \cdot \mathbf{J})^2 - \frac{1}{2}(\xi \eta_{xx} - \eta \xi_{xx})^2 - \frac{1}{2}(\xi \eta_{yy} - \eta \xi_{yy})^2$$
$$- (\xi \eta_{xy} - \eta \xi_{xy})^2, \tag{2.2.10}$$

$$\det \mathbf{M_{J}}_{\text{asym}} = (\xi_x \eta_y - \xi_y \eta_x)^2 = \omega^2. \tag{2.2.11}$$

$\mathbf{M_{J}}_{\text{asym}}$ contributes only to the vortices, since $\mathbf{M_{J}}_{\text{asym}}$ and circulations exchange direction (topological charge) under a reflection, and $\det \mathbf{M_{J}}_{\text{sym}}$ is zero at the vortex since $\xi = \eta = 0$ there. Conversely, $\mathbf{M_{J}}_{\text{sym}}$ only contributes to the extrema and saddles, being invariant under reflection, and $\det \mathbf{M_{J}}_{\text{asym}} = 0$ there by equation (2.2.4). The two kinds of zeros of $\mathbf{J}$ (critical points and phase singularities) can therefore be distinguished in equation (2.2.7), the number of circulations (the contribution from $\mathbf{M_{J}}_{\text{asym}}$) being the same as the number of phase singularities (equation (2.1.6)), the number of critical points $C_{\mathcal{A}}$ in area $\mathcal{A}$ being

$$C_{\mathcal{A}} = \int_{\mathcal{A}} \mathrm{d}^2 \mathbf{R} \delta^2(\mathbf{J}) |\det \mathbf{M_{J}}_{\text{sym}}|. \tag{2.2.12}$$

The number of saddles, $S_{\mathcal{A}}$ and extrema $E_{\mathcal{A}}$ are therefore

$$S_{\mathcal{A}} = \frac{1}{2}(Z_{\mathcal{A}} - I_{\mathcal{A}}) \tag{2.2.13}$$

$$E_{\mathcal{A}} = C_{\mathcal{A}} - S_{\mathcal{A}}. \tag{2.2.14}$$

Extrema are places where the divergence of current $\nabla \cdot \mathbf{J}$ is positive, since the sign of $\det \mathbf{M_{J}}_{\text{sym}}$ is the index of that point, and $\nabla \cdot \mathbf{J}$ is the only nonnegative term in equation (2.2.10). For solutions ψ_{H} of the planar Helmholtz equation (cf equation (1.5.2)),

$$\nabla_{\perp}^2 \psi_{\text{H}} + K^2 \psi_{\text{H}} = 0, \tag{2.2.15}$$

the current $\mathbf{J}_{\text{H}}$ is divergenceless since, by equation (2.2.1)

$$\nabla \cdot \mathbf{J}_{\text{H}} = \text{Im}(\nabla \psi_{\text{H}}^* \cdot \nabla \psi_{\text{H}} + \psi_{\text{H}}^* \nabla_{\perp}^2 \psi_{\text{H}}) = \text{Im}(-K^2 |\psi_{\text{H}}|^2) = 0. \tag{2.2.16}$$

In this case, the only critical points are divergenceless saddles (where $\text{tr} \, \mathbf{M_{J}}_{\text{sym}} = 0$), and in such waves the only mechanism for creation/annihilation of dislocations is that of (2.2.5).

There is a saddle near the edge dislocation in equation (1.2.5), on the selfintersection of the phase line labelled $\pi/2$ in figure (1.1), and the flow lines of current for this wave are shown in figure (2.2).
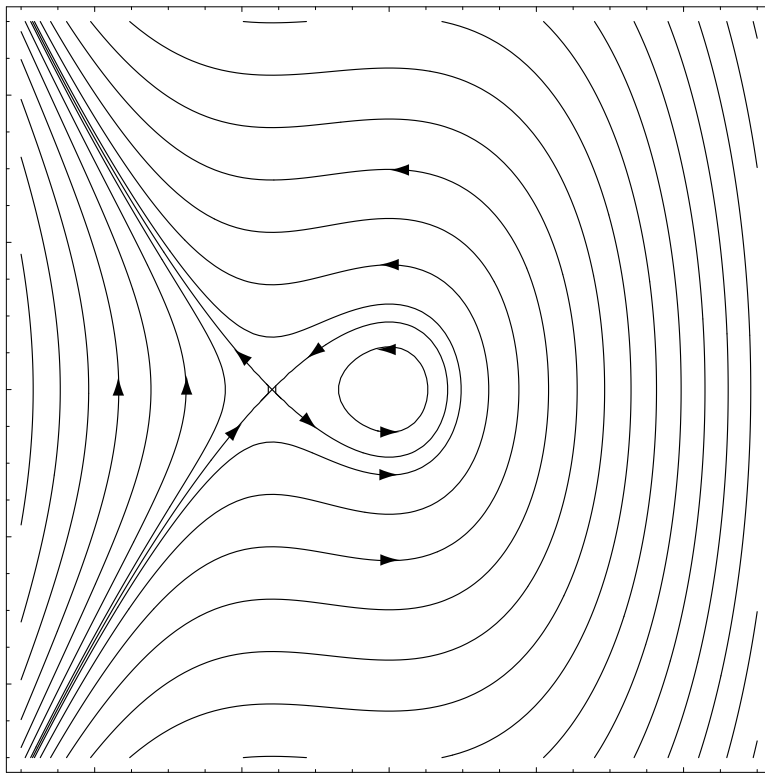
Figure 2.2: The flow of current around the edge dislocation (1.2.5), showing the circulation about the phase singularity and nearby saddle. (Figure courtesy of Michael Berry.)

2.3   Local structure of dislocation points

In this section we shall investigate the local structure of phase singularities, and see how the local intensity structure is related to the way the phase changes around the singularity, of possibly arbitrarily high strength. This is called the *dislocation core structure*. The core structure, as we are about to see, is elliptical, and many relevant properties of ellipses are discussed in appendix A.

In general, the intensity ρ^2 at the core is a quadratic form (where all derivatives and functions of them are taken at the singularity, translated to the origin),

$$
\begin{aligned}
\rho^2 &\approx \ |\mathbf{R} \cdot \nabla \psi|^2 \\
&= \ (\nabla \xi \cdot \mathbf{R})^2 + (\nabla \eta \cdot \mathbf{R})^2 \\
&= \ \mathbf{R}^T (\nabla \xi \otimes \nabla \xi + \nabla \eta \otimes \nabla \eta) \mathbf{R}.
\end{aligned}
\tag{2.3.1}
$$

The local contours of ρ^2 (and therefore of amplitude ρ) are therefore ellipses (see section A.2 for more details), defining the *core anisotropy ellipse*, discussed by [BD00, SS96] (also by [FF97], but the simple elliptical structure was missed). The equations for current (2.2.1) and its local expansion (2.2.3) give the local form for phase gradient

$$
\nabla \chi \approx \frac{\omega \wedge \mathbf{R}}{(\nabla \xi \cdot \mathbf{R})^2 + (\nabla \eta \cdot \mathbf{R})^2},
\tag{2.3.2}
$$

so, by equation (A.1.4), a polar plot of $\sqrt{|\nabla \chi|}$ around a circle centred on the dislocation is an ellipse with the same eccentricity as the ρ^2 contours (but with major and minor semiaxes interchanged). Equation (2.3.2) shows that $\nabla \chi$ must be in the $\mathbf{e}_\phi$ direction, therefore

$$
|\nabla \chi| = \frac{1}{R} \frac{\partial \chi}{\partial \phi}.
\tag{2.3.3}
$$

Taking the modulus of equation (2.2.2), (recalling that ω can be viewed as a vector perpendicular to the plane), $J(R) = \omega(0)R = \rho^2 |\nabla \chi|$, so

$$
R^2 \frac{\partial \phi}{\partial \chi} = \frac{\rho^2}{\omega(0)},
\tag{2.3.4}
$$

which is constant on the intensity contour, giving an analogue of conservation of angular momentum (defined as $R^2 \partial \phi / \partial \chi$). The interpretation is that equal area sectors of the core anisotropy ellipse are swept out in equal intervals of phase (this is related to the geometry of the auxiliary circle of the ellipse, see section A.1). The distribution of phase lines around the core ellipse is shown for two ellipses in figure (3.9). This version of Kepler's law is

identical to the conservation of angular momentum for a linear central force in classical mechanics, for instance the force experienced by a conical pendulum (see, for example, [Arn89]).

Of course, only for monochromatic waves is this quantity actually conserved with time, where the derivative in (2.3.4) with respect to χ can be replaced by t on the left hand side of equation (2.3.4), and each phase line (mod 2π) rotates about the ellipse once per cycle. This type of angular momentum is related to the quantum angular momentum of laser beams [APB99], as discussed by [SS96].

The eccentricity ε of the core ellipse can be found using the eigenvalues of the quadratic form in equation (2.3.1). The absolute squared length of $\nabla\psi$, is phase invariant, and shall be denoted

$$G \equiv |\nabla\psi|^2. \tag{2.3.5}$$

This is the unnormalised Stokes parameter S_0 (defined in (A.4.7)) for the core anisotropy ellipse, and the parameter S_3 is clearly equal to 2ω. The eccentricity, from equation (A.4.9), can be written in the simple gauge invariant form

$$\varepsilon = \frac{1}{\sqrt{2}\omega}(G^2 - 4\omega^2)^{1/4}\sqrt{G - \sqrt{G^2 - 4\omega^2}}. \tag{2.3.6}$$

We now consider the local structure of a high-order dislocation in a solution ψ of the (planar) Helmholtz equation, which we take to be of strength n ($n > 0$ for convenience) at the origin of the plane (as investigated in appendix A of [BD01a]). Obviously, for every phase, n phase contours must intersect at the origin. Around a circuit of the origin, the phase increases by $2n\pi$, implying that the first terms in an expansion of ψ around the origin are proportional to $\exp(\pm in\phi)$, and smoothness requires the R dependence to be proportional to R^n (since $R^n \exp(\pm in\phi) = (x \pm iy)^n$). This in turn implies that the leading term of $\nabla^2_\perp\psi$ goes as R^{n-2}, and therefore dominates $K^2\psi$ in the planar Helmholtz equation (2.2.15). Sufficiently close to the dislocation, the wave satisfies Laplace's equation $\nabla^2\psi = 0$, the most general solution of which, proportional to R^n, is

$$\begin{aligned}
\psi &= R^n(a_+ \exp(in\phi) + a_- \exp(-in\phi)) \\
&= a_+(x + iy)^n + a_-(x - iy)^n, \tag{2.3.7}
\end{aligned}$$

where $a_\pm$ are complex constants. In a circuit around the dislocation, the a_+ term winds $+n$ times around the origin of the ψ plane, the a_- term $-n$ times: in order for the phase to increase in the correct sense, $|a_+| > |a_-|$ (and vice versa if $n < 0$).

The solution (2.3.7) can also be demonstrated by requiring that ξ and η independently satisfy Laplace's equation, vanishing to order n, ie

$$\begin{aligned}
\xi &= R^n(\alpha_1 \cos n\phi + \alpha_2 \sin n\phi) \\
\eta &= R^n(\beta_1 \cos n\phi + \beta_2 \sin n\phi)
\end{aligned} \tag{2.3.8}$$

where $\alpha_1, \alpha_2, \beta_1, \beta_2$ are real and easily related to $a_\pm$. This local structure is more restrictive than for an high-order phase singularity in an arbitrary scalar field, made up of terms in x and y like $x^j y^{n-j}$, $(0 \leq j \leq n)$, caused by the reduction near the zero from the Helmholtz to the Laplace equation. If the wave is nonmonochromatic, but has Fourier components satisfying the Helmholtz equation (as discussed in section 1.5), then the above still holds (since sufficiently close to the dislocation, the laplacian of each component still dominates the $K^2 \psi$ term). The contours of intensity ρ^2, by equation (2.3.7), lie on the parametric curve

$$R(\phi) = \sqrt[n]{\frac{\rho}{|a_+ \exp(in\phi) + a_- \exp(-in\phi)|}}, \tag{2.3.9}$$

whose nth power is one of the generalised ellipses of equation (A.1.7) and figure (A.2), for n integer. After setting n, there are only four real (two complex) parameters describing the figure, the same as for a usual ellipse. [FB00, Fre01] have made a study of the behaviour of higher index phase critical points in two dimensional wavefields, which we do not discuss here. The angular momentum phase relation of equation (2.3.4) clearly generalises to the higher order structures. The above construction also applies to high-order dislocation lines in space (satisfying the three dimensional Helmholtz equation), the appropriate details being given in [BD01a] appendix A.

Topologically, a strength n dislocation ($n > 0$ for convenience) not only requires the coalescence of n strength 1 dislocations, but also $n-1$ saddles, in order that the Poincaré index of the high strength singularity be 1 (the phase contours still radiate outwards, so the current, and $\nabla \chi$, circulate with index 1). This requires the $n(n+1)$ conditions

$$\begin{aligned}
\partial_x^q \partial_y^{p-q} \xi &= 0, \quad 0 \leq q \leq p, \quad 0 \leq p \leq n-1 \\
\partial_x^q \partial_y^{n-q} \xi &\neq 0, \quad 0 \leq q \leq n, \\
\partial_x^q \partial_y^{p-q} \eta &= 0, \quad 0 \leq q \leq p, \quad 0 \leq p \leq n-1 \\
\partial_x^q \partial_y^{n-q} \eta &\neq 0, \quad 0 \leq q \leq n,
\end{aligned} \tag{2.3.10}$$

to be satisfied at the dislocation, implying that such a dislocation has a higher codimension than that required from simply superposing n dislocations. The behaviour of the unfolding

of high strength dislocations is used to elucidate the structure of knotted phase singularities in chapter 5; what happens if the $\neq$ conditions in (2.3.10) are violated is also explored there.

As mentioned in section 1.2, much of the initial study of wave dislocations was by way of comparison with crystal dislocations; in particular, [NB74] define the Burgers vector of a dislocation to be the wave propagation direction if, apart from the dislocation, the wavefronts are otherwise well ordered like a plane wave. [Nye81] discusses in some detail a certain complex procedure for constructing a Burgers vector for an arbitrary wave dislocation. In the two dimensional waves considered here, all dislocations are of edge type, but it is not clear whether there is a natural topological definition of the Burgers vector for dislocations in arbitrary waves; it is likely that the appearance of the Burgers vector in waves like the edge dislocated one in equation (1.2.5) is due to the fact that there is an overall exponential factor representing a plane wave propagating tin the y direction. We shall not consider the Burgers vector for edge dislocations further.

Although the anisotropy ellipse of the edge dislocation in equation (1.2.5) is a non-generic circle, it can be easily modified to have elliptic structure:

$$\psi = (x/a + \mathrm{i}y/b)\exp(\mathrm{i}ky), \tag{2.3.11}$$

where a, b are real and positive, and the ellipse semiaxes are a, b the x, y directions, with the major semiaxis in the x direction if $a > b$.

2.4 Motion and velocity of dislocations

We now consider ψ as a time dependent solution of some wave equation in two or three dimensions (where position is denoted by $\mathbf{r}$. We have already seen in section 1.5 that the dislocation pattern is stationary for monochromatic waves. This is one of the main reasons that optical dislocations are significant, because, in a monochromatic field (even a random one, such as a laser speckle pattern), the dislocations are stationary although the frequencies are very high (and time-dependent features move too quickly to register).

If the waves are not monochromatic (but still complex), the dislocations may move, and, as we shall see, may move arbitrarily quickly (and independently of the speed c of the dispersionless Fourier components of the wave). At first this may appear a contradiction of relativity, but since dislocations are morphologies of the field (arising from the geometry of

the interference pattern), they are not subject to the laws which apply to physical objects. Since they are zeros, they carry no energy, and cannot be used to transmit information. The speed of other morphological objects is similarly unrestricted: consider the speed of the intersection of scissor blades (which accelerates increasingly if the scissors are opened at a constant rate), or the end of a searchlight beam (which, for a sufficiently strong beam, also moves arbitrarily fast, arbitrarily far away).

A simple example of a wave with a single moving dislocation is a modification of the edge dislocated wave (1.2.5), given by [NB74] equation (42). The wave now satisfies the time dependent wave equation, and is

$$\psi = (ax - kx^2 + \mathrm{i}(b(y - ct) - y)) \exp(\mathrm{i}(ky - \omega_k t)), \qquad (2.4.1)$$

where $c = \omega_k/k$, the speed of the wavefronts, and a, b are arbitrary and dimensionless. There are two dislocations, with x coordinates $0, a/k$ and common y coordinate $bct/(b-1)$; they therefore are moving in the y-direction with speed $bc/(b-1)$, which can be arbitrarily fast for b arbitrarily close to 1. If the wave admits the definition of a Burgers vector, then it is possible for the dislocation to glide and climb, as shown by [NB74].

We now derive a general expression (valid in two or three dimensions) for the velocity of a dislocation, which shall be useful later on. The dislocation conditions ($\xi(\mathbf{r}, t) = 0, \eta(\mathbf{r}, t) = 0$) mean that, on dislocations, the total derivative is also zero:

$$\nabla\psi \cdot \mathrm{d}\mathbf{R} + \psi_t \mathrm{d}t = 0. \qquad (2.4.2)$$

The dislocation velocity $\mathbf{v}$ is defined $\mathbf{v} = \mathrm{d}\mathbf{r}/\mathrm{d}t$, and so by equation (2.4.2), satisfies the simultaneous equations

$$\mathbf{v} \cdot \nabla\xi = -\xi_t, \qquad \mathbf{v} \cdot \nabla\eta = -\eta_t \qquad (2.4.3)$$

which are easily solved to give (where $\boldsymbol{\omega}$ is perpendicular to the (x, y) plane)

$$\mathbf{v} = \frac{1}{\omega^2}(\eta_t \nabla\xi - \xi_t \nabla\eta) \wedge \boldsymbol{\omega}. \qquad (2.4.4)$$

It is easy to verify that a monochromatic wave, with a single time component gives zero in (2.4.4). The velocity squared is

$$v^2 = \frac{1}{\omega^2}(\xi_t^2(\nabla\eta)^2 + \eta_t^2(\nabla\xi)^2 - 2\xi_t\eta_t\nabla\xi \cdot \nabla\eta). \qquad (2.4.5)$$

2.5   Nodal lines in three dimensional space

One of the strengths of the analogy between phase singularities and crystal dislocations is that it holds in three dimensional space as well as two, and, as we have seen, the codimension 2 nature of phase singularities imply that they are generically lines in space. This fact was emphasised by [NB74], but has only recently been appreciated by the experimental optics community [Fre00]. Many of the results and formulae that were derived for dislocations in two dimensions also hold in three, as we shall see. In the following, vectors are assumed three dimensional (including the ∇ operator) and position in space is given by the vector $\mathbf{r} = (x, y, z)$.

The real and imaginary parts ξ, η of the three dimensional complex field ψ have their zero level sets (contours) generically on surfaces (each is codimension 1), which intersect along lines of phase singularity. The geometry of dislocation lines is therefore the geometry of the generally curved intersections of the family of surfaces $\operatorname{Re} \psi \exp(-\mathrm{i}\chi) = 0$ for each $\chi \bmod \pi$. These surfaces are, of course, orientable (twosided) since they separate regions where the real function is positive and negative. The surfaces may be infinite or finite (and closed), and may (nongenerically) self intersect; the dislocation lines may also have any of these properties (they may be infinitely long or closed loops, possibly crossing each other).

Dislocation strength, as the change of phase (divided by 2π) in a circuit C enclosing the dislocation line, is identically defined as for the two dimensional case (1.2.4). The dislocation line cannot end (the Dirac string of section 1.6 is an anomaly, because the quantum wavefunction does not have a well-defined global phase), and the dislocation strength is conserved as C slides up and down the dislocation line. This endows a direction along the dislocation, that about which the phase increases in a right-handed sense: this defines the *topological current*, which, like topological charge, is a signed integer (the sign always being positive along the direction of the topological current). The topological current is not to be confused with the field current $\mathbf{J}$, or its three dimensional analogue $\mathbf{j}$. The topological current is generically 1, with higher currents only when the dislocation corresponds to self-intersection of the phase surfaces.

A convenient vector which points in the direction of topological current is the three-dimensional generalisation $\boldsymbol{\omega}$ of the vorticity, defined as in (2.2.2), as the vector product $\nabla \xi \wedge \nabla \eta$, which is not now restricted to the z-direction. Since $\nabla \xi, \nabla \eta$ are the normals to

the contour surfaces, their vector product must lie in each surface, and the right-handed sense of the vector product follows the direction of topological current. The dislocation length $\ell_{\mathcal{V}}$, therefore, in a volume $\mathcal{V}$, is the integral

$$\ell_{\mathcal{V}} = \int_{\mathcal{V}} \mathrm{d}^3\mathbf{r}\, \delta(\xi)\delta(\eta)|\nabla\xi \wedge \nabla\eta|. \tag{2.5.1}$$

This generalises (2.1.6); the δ-functions again restrict the integral to the intersection of the zeros of ξ, η, but now the jacobian is the length of a vector; locally, $\nabla\xi, \nabla\eta$ are perpendicular to the dislocation, and give the same jacobian as in the two-dimensional case in this plane.

The 3-current $\mathbf{j}$ is also defined in an identical way to its two-dimensional analogue (2.2.1) (where ∇ in the definition is now three-dimensional). $\boldsymbol{\omega}$ is half its curl, and $\mathbf{j}$ vanishes on dislocation lines. The other current zeros are at critical points of $\nabla\chi$, which, being a standard three-dimensional gradient field, only vanishes at points. These critical points of phase (and zeros of current) have more complicated indices than their two-dimensional analogues; the local matrix (cf (1.2.8)) is now three-dimensional, with three nonzero eigenvalues, which may be three positive (a minimum), three negative (a maximum) or two positive/negative and one negative/positive. They correspond to the appearance (or disappearance) of a phase surface at a point, and in solutions of the (three-dimensional) Helmholtz equation, must be divergenceless (as in (2.2.16)). The dislocations are no longer on an equal footing with the phase critical points in three dimensions (the dimensions are different), and we shall not consider three-dimensional phase critical points further.

$\nabla\xi, \nabla\eta$ are perpendicular to the dislocation, and in this plane normal to the dislocation, the local two-dimensional structure of dislocations in section 2.3 applies in three dimensions, including the structure for higher strength. The core ellipse is defined in the $(\nabla\xi, \nabla\eta)$ plane, and has eccentricity (2.3.6), with appropriate generalisation of (2.3.5). The velocity formulae (2.4.4), (2.4.5) similarly hold; the motion again is only in the $(\nabla\xi, \nabla\eta)$ plane; since the dislocation is not a physical object, but a morphology, the only motion it can sensibly have is transverse motion.

As before, the only independent phase surfaces that need to be considered are those of ξ and η, the others being made up from them by (2.1.2). Dislocation loops can be created or annihilated (as a parameter such as time is varied) when one phase surface passes through another; at the moment of creation, the two surfaces touch and the gradients

a

b
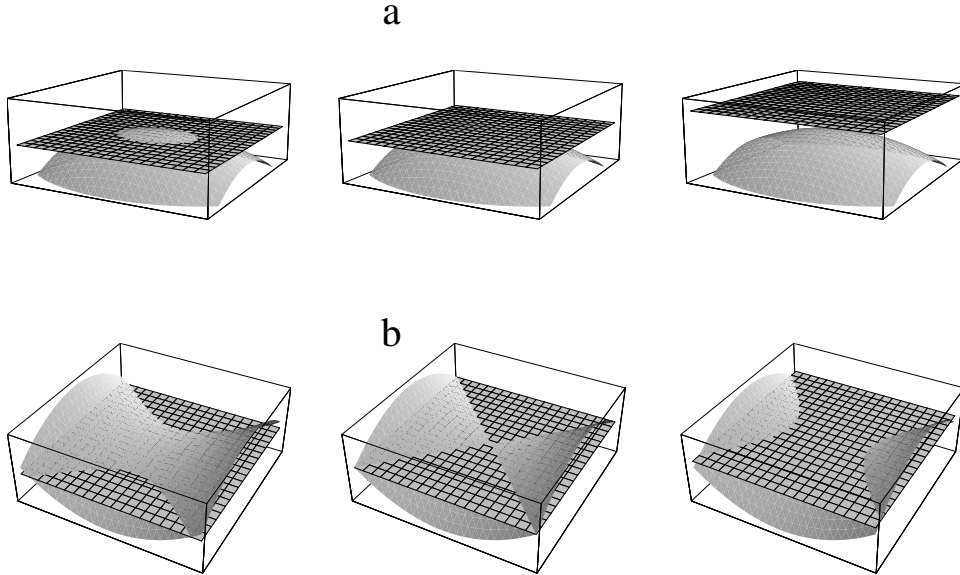
Figure 2.3: Changing the topology of dislocation lines. (a) a loop is created/destroyed as one surface touches another; (b) the dislocation reconnects as one surface passes through a saddle point in the other (see also figure (2.5)).

coincide ($\nabla\xi \wedge \nabla\eta = 0$). A simple example of this, shown in figure (2.3a), occurs when a function with a maximum passes through a plane. Dislocation lines reconnect when one of the contour surfaces passes through a saddle point in the other (2.3b), also where the directions of the two normals coincide, and is discussed in more detail in the following section.

Mathematically, the field ψ provides two kinds of foliation of space, one on which the 'leaves' are one-dimensional, the other two-dimensional. The latter is simply the decomposition of space into its individual phase surfaces; this foliation breaks down at the phase singularity, where all of the different phase surfaces intersect, and also when $\boldsymbol{\omega} = \mathbf{0}$ (which is where loops are created/destroyed, as in figure (2.3)). The other foliation is along the lines labelled by the complex number $\xi + i\eta$ (the dislocations being the only choice of values invariant under a global phase change).

We conclude the section with a brief examination of how the spatial structures in three-dimensional space (x, y, z) are related to the two-dimensional ones considered above (for convenience, in the (x, y) plane). Firstly, the dislocation lines cross the (x, y) plane in points, the topological charge of which depends on whether the topological current is
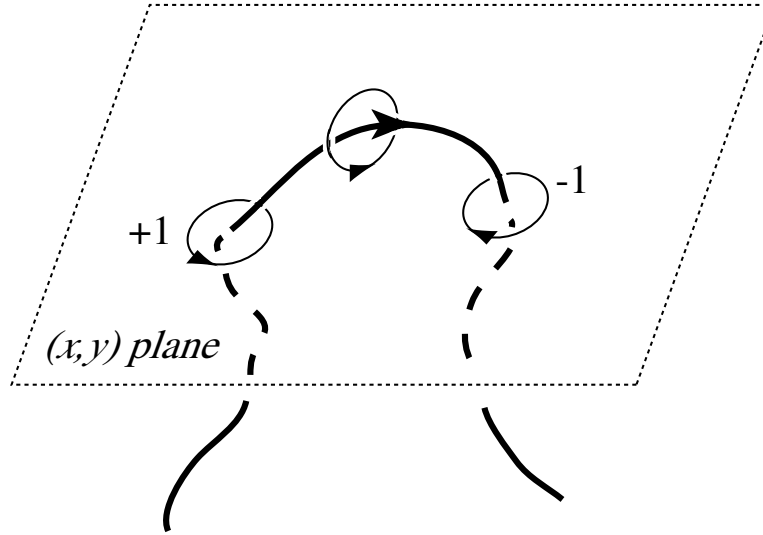
Figure 2.4: Indicating the underlying three dimensional geometry of dislocation creation and annihilation for plane sections of spatial fields

directed into the plane ($\omega_z < 0$), in which case the apparent topological charge of the point is negative, or out of the plane ($\omega_z > 0$), in which case the apparent charge is positive (generically, $\omega_z \neq 0$). Therefore, if two dislocation points are created/annihilated in a plane section of a three dimensional field, the three-dimensional geometry is just that of an arched dislocation passing through the plane, as in figure (2.4). The perceived core ellipse is that of a cylinder with elliptic cross section being obliquely sliced: the eccentricity in the plane section is higher than transverse to the dislocation. If the dislocation is moving, the motion in the plane is faster than transverse to the line in space, because the position of the intersection point is moving along the dislocation (it is a morphology of a morphology). The planar dislocation velocity acquires a virtual component along its direction.

The 2-current $\mathbf{J} = (j_x, j_y)$, and is zero when $\mathbf{j}$ is in the z-direction alone, that is when a phase surface (with normal $\nabla\chi$) touches the (x, y) plane. It is a maximum, minimum, or saddle according to the local geometry of the surface, the geometry being the same as in figure (2.3), but the plane not necessarily being a phase surface.

The geometry and topology of dislocation lines is obviously richer than dislocation points, since the lines may touch and reconnect, be curved, and have their phase cores twisting around them. These possibilities are investigated in the following sections.
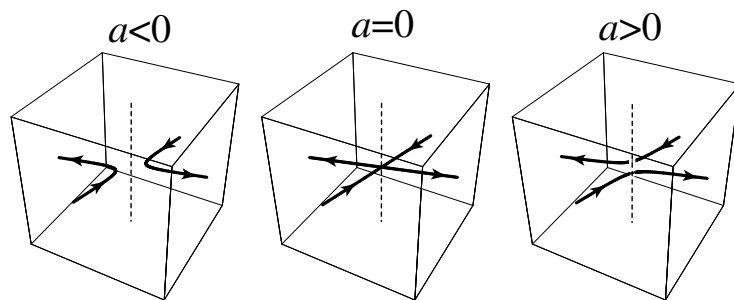
Figure 2.5: Dislocation reconnection, described by equation (2.6.1) with $\beta = \pi/12$. Arrows give the direction of topological current, and the dashed line shows where $\boldsymbol{\omega} = \mathbf{0}$. The phase surfaces for these dislocations are those of figure (2.3b).

2.6   Dislocation reconnection

An important topological question of dislocation lines in space is whether they can cross, and if so, do the lines cross inertly or reconnect (preserving the sense of the topological current). One would expect such events to have codimension 4, and we shall see that this is indeed the case, albeit in surprising circumstances. These changes take place as a parameter a varies, which may represent time evolution. A simple example of a topological reconnection as a varies is the crossing of a saddle, as in figure (2.3b); the corresponding picture for the dislocations in this case is shown in figure (2.5). The phase singularities are two hyperbolae perpendicular to z, crossing and reconnecting at the origin. A solution of the Helmholtz equation (1.5.2) (with k set to 1) with phase contours that behave in this way is the wave

$$\psi = \left(\frac{xy}{\sin 2\beta} + \frac{1}{2}(x^2 + y^2) + a + \mathrm{i}z \right) \exp(\mathrm{i}z), \qquad (2.6.1)$$

where β is any angle between (but not including) 0 and $\pi/2$, (noting the correction from [BD01b] equation (4)). The dislocations of this function are plotted in figure (2.5). They have asymptotes in the directions (ϕ azimuthal angle) $\phi = \beta, \pi/2 - \beta, \pi + \beta, 3\pi/2 - \beta$. This appears to be a normal form for dislocation reconnection.

When dislocations cross, the vorticity $\boldsymbol{\omega}$ vanishes, since there is no unique dislocation direction (in equation (2.6.1), the dislocations point in opposite directions, as is clear from figure (2.5)). Now, $\boldsymbol{\omega}$, as the vector product of two real vector fields, has zeros of codimension 2 ('L lines'), as discussed fully in chapter 4 (they are, in fact, L lines of $\nabla \psi$). Dislocation reconnection appears to occur generically whenever a dislocation line

encounters an L line. When a dislocation loop is born or disappears (as in figure (2.3a)), the phase surfaces touch in a similar way, and there is an L line threading the dislocation loop. Therefore, the L lines of $\nabla\psi$ play an important role in the topology of dislocation lines, a point that shall be returned to in chapter 5. In equation (2.6.1), the L line ($\boldsymbol{\omega} = \mathbf{0}$) is along the z-axis, depicted by the dotted line in figure (2.5).

Dislocations may be made to cross inertly, without reconnecting. A function with phase singularities doing just this is easy to construct by multiplying two functions, each with one of the desired zero lines, such as

$$\psi = (x + \mathrm{i}z)(y + \mathrm{i}(z - a)), \tag{2.6.2}$$

with a stationary dislocation along the y-axis from the first factor, the second providing a dislocation in the x-direction, moving along the z-direction with $y = 0$; the two cross when $a = 0$. However, this function is unstable against perturbation, and the addition of a plane wave in any direction destroys this benign crossing, the reaction instead taking place by two reconnections. Moreover, equation (2.6.2) is not a solution to any wave equation described in section 1.5, and the addition of extra terms in order to satisfy a wave equation destroys the crossing. The ξ, η contours of equation (2.6.2) are quite singular as a approaches 0, and it is likely that dislocations can only cross inertly (that is, without a topological reconnection) if the phase surfaces do change their topology, which is, of course, not generic.

[BBBBS00] show a wave similar to equation (2.6.2) which satisfies the time-dependent three- dimensional Schrödinger equation, but in fact, a detailed examination once again shows that the interaction takes place via a reconnection of the type (2.6.1). This is therefore a difference between the quantised vortices of phase singularities and the vortices of hydrodynamics, which may not cross [Mof69]. This is used in chapter 5 to tie and untie knotted dislocation lines.

2.7   Curvature and torsion of dislocation lines

Generically, dislocation lines are curved (since not even do the contour surfaces $\xi = 0, \eta = 0$ usually contain straight lines), and nonplanar, that is, they have nonvanishing curvature κ and torsion τ. A brief discussion of the geometry of space curves is given here; details may be found in standard texts on classical differential geometry (such as [dC76, Eis60,

Gra93]). The differential geometry of curves formed by the intersection of two surfaces is an interesting and general problem in classical differential geometry (an example of such a curve is Viviani's curve, which is the intersection of a sphere and a cylinder [Gra93]).

Generically, dislocation lines are regular, since any inflexion points can easily be perturbed away, so the Frenet frame is defined everywhere on the curve. As discussed in section 2.5, the vorticity $\boldsymbol{\omega}$ is in the direction of the curve, so the tangent vector is

$$\mathbf{T} = \boldsymbol{\omega}/\omega, \tag{2.7.1}$$

and the orientation of the singularity curve is that of the topological current. When $\omega(\mathbf{r}) = 0$ dislocation lines cross, which is not generic enough to concern us here.

Along the curve, ω is constantly changing ($\omega' \neq 0$), so provides an acceptable non-arclength parameterisation of the dislocation curve; the arclength derivative along the curve can be written is $(\boldsymbol{\omega}/\omega \cdot \nabla)$, and is denoted $\bullet'$. The *curvature* κ of the curve is a measure of the deviation of $\mathbf{T}$ from $\mathbf{T}$, defined by

$$\kappa = |\mathbf{T}' \cdot \mathbf{T}|, \tag{2.7.2}$$

which generically does not vanish. The direction of $\mathbf{T}'$ is called the *normal direction* $\mathbf{N}$, and locally the curve lies in the $(\mathbf{T}, \mathbf{N})$ (osculating) plane. The osculating plane normal $\mathbf{T} \wedge \mathbf{N}$ is called the *binormal* $\mathbf{B}$, and the orthogonal frame $(\mathbf{T}, \mathbf{N}, \mathbf{B})$ is called the *Frenet frame* of the curve.

In terms of ω, the curvature (2.7.2) is written

$$\kappa = \frac{|\boldsymbol{\omega} \wedge (\boldsymbol{\omega} \cdot \nabla)\boldsymbol{\omega}|}{\omega^3}, \tag{2.7.3}$$

where quantities are evaluated on the dislocation; this expression is phase invariant.

This formula is considerably simplified if coordinates are chosen such that the dislocation passes through the origin when $\boldsymbol{\omega}$ lies in the z-direction, so the arc length derivative is just ∂_z. In this case, where $\mathbf{e}_z$ is the unit vector in the z-direction,

$$
\begin{aligned}
\kappa^2 &= \frac{(\mathbf{e}_z \wedge \partial_z \boldsymbol{\omega})^2}{\omega^2} \\
&= \frac{(\partial_z \boldsymbol{\omega})^2 - (\partial_z \omega_z)^2}{\omega^2} \\
&= \frac{\xi_{zz}^2 (\nabla \eta)^2 + \eta_{zz}(\nabla \xi)^2 - 2\xi_{zz}\eta_{zz}\nabla \xi \cdot \nabla \eta}{\omega^2}
\end{aligned}
\tag{2.7.4}
$$

which is formally the same as the expression (2.4.5) for the squared velocity of a dislocation line, with ξ_{zz}, η_{zz} for ξ_t, η_t. This is not surprising, since the rate of change of the surfaces

ξ, η at the origin in the z-direction is proportional to their second derivatives with respect to z, provided $\boldsymbol{\omega} \| \mathbf{e}_z$ at the origin.

The *torsion* τ of a curve is a measure of the departure from planarity of the curve, and is defined

$$\tau = \mathbf{N}' \cdot \mathbf{B}, \tag{2.7.5}$$

which is clearly 0 for a plane curve. In terms of ω, it is

$$\tau = \frac{\boldsymbol{\omega} \wedge (\boldsymbol{\omega} \cdot \nabla)\boldsymbol{\omega} \cdot (\boldsymbol{\omega} \cdot \nabla)^2 \boldsymbol{\omega}}{\kappa^2}, \tag{2.7.6}$$

again phase invariant. (2.7.6) simplifies, but not significantly, choice of coordinates used for (2.7.4).

A helix is defined as a curve with constant (nonzero) curvature and torsion. If it has radius a and pitch (distance between successive coils) $2\pi b$, the curvature and torsion are

$$\kappa = \frac{a}{a^2 + b^2}, \qquad \tau = \frac{b}{a^2 + b^2}. \tag{2.7.7}$$

Although κ, τ are not invariant if the curve is rescaled, their ratio a/b is. In the limit of the curve approaching a straight line, (for which $\kappa = 0$ and $\mathbf{N}$ is not defined), where $a \to 0, \kappa \to 0$ and b constant, the torsion is approaches the constant value $1/b$. Torsion is not an entirely useful geometric measure.

As Nye discusses ([Nye99], pp103-104), a perturbation of the screw dislocation (2.8.1) by a plane wave ψ_ε in the z-direction

$$\psi_\varepsilon = -\varepsilon \exp(\mathrm{i}k_1 z), \tag{2.7.8}$$

so, after perturbation, the real and imaginary parts being zero give the equations

$$\begin{aligned} x\cos kz - y\sin z &= \varepsilon \cos k_1 z \\ x\sin kz + y\cos z &= \varepsilon \sin k_1 z, \end{aligned} \tag{2.7.9}$$

with curvature and torsion

$$\kappa = \frac{\varepsilon(k_1 - k)^2}{1 + \varepsilon^2(k_1 - k)^2}, \qquad \tau = \frac{k_1 - k}{1 + \varepsilon^2(k_1 - k)^2}, \tag{2.7.10}$$

from equations (2.7.7).

2.8 Twist and twirl

Much of the geometric interest of dislocation lines is not the geometry of the line itself, but the geometry of the phase structure which the dislocation organises. In general, a given phase surface in the neighbourhood of a dislocation is a helicoid, the simplest example being the example of [NB74],

$$\psi = (x + \mathrm{i}y)\exp(\mathrm{i}kz) = R\exp(\mathrm{i}\phi)\exp(\mathrm{i}kz), \qquad (2.8.1)$$

the helicoidal geometry of the phase surfaces being reminiscent of a crystal screw dislocation; by the crystal-wave analogy, this wave dislocation shall be referred to as a screw dislocation. Each of the phase surfaces is a helicoid with pitch $2\pi/k$, left handed if $k > 0$. The phase surfaces are all identical, and rotation about the z-axis by ϕ_0 is equivalent to multiplying the function by a phase $\exp(\mathrm{i}\phi_0)$; such functions (considered at length in chapter 5) shall be referred to as *cylindrically phase-symmetric*.

In crystals, a screw dislocation has its Burgers vector in the dislocation tangent direction, unlike an edge dislocation, where the Burgers vector is perpendicular to the dislocation direction (consider, for example, the wave edge dislocation (1.2.5) trivially as a function of z as well). The problem of writing a general Burgers vector, discussed above for edge dislocations, generalises to screw dislocations (and mixed edge-screw dislocations, discussed by [Nye99, Nye81]). The Burgers vector formalism has been used to analyse the dislocation structure of quantum mechanical waves [Hol87], but the formalism of this work is not appropriate at our level of discussion. Although our terminology used may imply the presence of a Burgers vector, we shall not derive an expression for it, instead concentrating on geometric features of the phase surfaces.

Nevertheless, it is still desirable to quantify the degree to which a dislocation line is twisted, that is, the rate of rotation of the phase helicoids with respect to distance along the line (ie, with respect to direction parallel transported along the line) . For a given phase surface $u = u_\chi = 0$ (see 2.1.2), the rate of rotation $Tw(\chi)$ is easily calculated to be the arc length derivative of the azimuthal angle of $\nabla u = \mathbf{U}$ in local cylindrical coordinates. This is the instantaneous twist of the local phase ribbon, as might be used in a Calugareanu-type calculation (see, for instance, [Ada94, Han98a]). For the remainder of this section the dislocation, when convenient, shall be assumed to cross the origin with $\mathbf{T}(\mathbf{0}) = \mathbf{e}_z$, and cartesian and cylindrical coordinates are used interchangeably. Also, $\mathbf{X}, \mathbf{Y}$ are used to denote $\nabla\xi, \nabla\eta$ respectively throughout this section. Thus, at the origin, the

local rate of rotation of $\mathbf{U}$ is

$$
\begin{aligned}
Tw(\chi) &= (\phi_z)_\chi \\
&= \partial_z \arctan \frac{u_y}{u_x} \\
&= \frac{u_x u_{yz} - u_y u_{xz}}{u_x^2 + u_y^2} \\
&= \frac{\mathbf{T} \cdot (\mathbf{U} \wedge \mathbf{U}')}{U^2} \\
&= \frac{\left(\begin{array}{c} \mathbf{T} \cdot \left((\mathbf{X} \wedge \mathbf{X}') \cos^2 \chi + (\mathbf{Y} \wedge \mathbf{Y}') \sin^2 \chi \right. \\ \left. +(\mathbf{X} \wedge \mathbf{Y}' + \mathbf{Y} \wedge \mathbf{X}') \cos \chi \sin \chi \right) \end{array}\right)}{X^2 \cos^2 \chi + Y^2 \sin^2 \chi + 2\mathbf{X} \cdot \mathbf{Y} \sin \chi \cos \chi}
\end{aligned}
\tag{2.8.2}
$$

where the last two equalities holds generally, and $\bullet'$ represents the arc length derivative as before. Applying this formula to the screw dislocated wave (2.8.1) gives $-k$ for every choice of χ, as one would expect (the sign indicates the handedness of the helicoid in the dislocation direction). However, if the core ellipse is not circular, there are problems; in particular, if the screw dislocated wave (2.8.1) is modified using (2.3.11),

$$
\psi = (x/a + \mathrm{i}y/b) \exp(\mathrm{i}kz),
\tag{2.8.3}
$$

with a, b real and positive, $a > b$ for simplicity. The cross section of the dislocation core is now an ellipse, with semiaxes in the x, y directions. At the origin, $Tw(\chi)$ is now

$$
Tw(\chi) = \frac{-kab}{b^2 \cos^2 \chi + a^2 \sin^2 \chi},
\tag{2.8.4}
$$

and depends on the phase χ chosen. To get a measure of Tw for the whole dislocation, it is necessary to average over all phase surfaces; in the following this is done by averaging with respect to phase χ.

If χ is measured from the rectifying phase χ_0 of (A.2.6) ($\chi \to \chi + \chi_0$, so $\mathbf{X}, \mathbf{Y}$ become $\mathbf{X}_0, \mathbf{Y}_0$ with $\mathbf{X}_0 \cdot \mathbf{Y}_0 = 0$), the last line of (2.8.2) is simplified,

$$
Tw(\chi) = \frac{\mathbf{T} \cdot ((\mathbf{X}_0 \wedge \mathbf{X}_0') \cos^2 \chi + (\mathbf{Y}_0 \wedge \mathbf{Y}_0') \sin^2 \chi + (\mathbf{X}_0 \wedge \mathbf{Y}_0' + \mathbf{Y}_0 \wedge \mathbf{X}_0') \cos \chi \sin \chi)}{X_0^2 \cos^2 \chi + Y_0^2 \sin^2 \chi},
\tag{2.8.5}
$$

and now $Tw(\chi)$ may be integrated with respect to χ, giving the *phase-averaged twist* Tw_χ,

$$
\begin{aligned}
Tw_\chi &= \frac{1}{2\pi} \int_0^{2\pi} \mathrm{d}\chi \, Tw(\chi) \\
&= \frac{\mathbf{T} \cdot (Y_0(\mathbf{X}_0 \wedge \mathbf{X}_0') - X_0(\mathbf{Y}_0 \wedge \mathbf{Y}_0'))}{X_0 Y_0 (X_0 + Y_0)},
\end{aligned}
\tag{2.8.6}
$$

which gives $-k$ in the case of (2.8.3). This is not the only way the twist can be averaged; $Tw(\chi)$ can be averaged with respect to ϕ instead; using the angular momentum relation (2.3.4), $\partial\phi/\partial\chi = \rho^2/R^2\omega$, and rewriting the $Tw(\chi)$ denominator using the auxiliary circle ellipse relation (2.3.1) at phase χ, it is possible to derive the *azimuth-averaged twist* Tw_ϕ, in the limit $R \to 0$,

$$
\begin{aligned}
Tw_\phi &= \frac{1}{2\pi}\int_0^{2\pi} \mathrm{d}\phi\, w(\chi) \\
&= \frac{1}{2\pi}\int_0^{2\pi} \mathrm{d}\chi\, \frac{\partial\phi}{\partial\chi} w(\chi) \\
&= \frac{1}{2\pi}\int_0^{2\pi} \mathrm{d}\chi\, \frac{\rho^2}{R^2\omega} \\
&\qquad \times \frac{\mathbf{T}\cdot((\mathbf{X}\wedge\mathbf{X}')\cos^2\chi + (\mathbf{Y}\wedge\mathbf{Y}')\sin^2\chi + (\mathbf{X}\wedge\mathbf{Y}' + \mathbf{Y}\wedge\mathbf{X}')\cos\chi\sin\chi)}{\rho^2(\chi)/R^2} \\
&= \frac{\mathbf{T}\cdot(\mathbf{X}\wedge\mathbf{X}' + \mathbf{Y}\wedge\mathbf{Y}'\sin^2\chi)}{2\omega} \\
&= \frac{\mathrm{Re}\{\mathbf{T}\cdot\nabla\psi^*\wedge\nabla\psi'\}}{2\omega}.
\end{aligned}
\tag{2.8.7}
$$

This also is $-k$ in the example, but is clearly different in form from the phase-averaged (2.8.6).

In [Ber01a], Berry instead works with $(\chi_z)_{\phi\,\mathrm{const}}$, measuring the rate of change of phase with respect to a parallel transported position ϕ along the dislocation, and averages in a strange way around ϕ (in the limit $R \to 0$) to get the *screwness* σ, (reproduced from [Ber01a] equations (6),(7))

$$
\begin{aligned}
\sigma &\equiv \lim_{R\to 0} \frac{\int_0^{2\pi} \mathrm{d}\phi\, \rho^2(\mathbf{R})\chi'(\mathbf{R})}{\int_0^{2\pi} \mathrm{d}\phi\, \rho^2(\mathbf{R})} \\
&= \lim_{R\to 0} \frac{\int_0^{2\pi} \mathrm{d}\phi\, j_z(\mathbf{R})}{\int_0^{2\pi} \mathrm{d}\phi\, \rho^2(\mathbf{R})} \\
&= \lim_{R\to 0} \frac{\int_0^{2\pi} \mathrm{d}\phi\, (\xi(\mathbf{R})\eta'(\mathbf{R}) - \eta(\mathbf{R})\xi'(\mathbf{R}))}{\int_0^{2\pi} \mathrm{d}\phi\, (\xi^2(\mathbf{R}) + \eta^2(\mathbf{R}))} \\
&= \frac{\mathbf{X}\cdot\mathbf{Y}' - \mathbf{Y}\cdot\mathbf{X}'}{X^2 + Y^2} \\
&= \frac{\mathrm{Im}\,\nabla\psi^*\cdot\nabla\psi'}{G}
\end{aligned}
\tag{2.8.8}
$$

where (2.2.1) has been used in the second equality, and a Taylor expansion in the fourth. This gives $-k$ in the simple example (2.8.3). [1] It is also possible to average the twist $Tw(\chi)$

---

[1] The derivation of (2.8.8) was heuristic to get a formula that gives this correct answer.

of the helicoids with respect to ϕ, or the rate of change of phase at constant azimuth χ' with respect to χ; the ones listed here (Tw_χ and σ) give the answer $-k$ for (2.8.3); they obviously measure subtly different properties when the situation is more complicated, but it is not clear what these are at this level of analysis.

The problem would seem to originate from the fact that the phase ellipse is usually rotating as well, and since this organises the transverse phase lines, the changing pitches of the different change at different rates. We define the rotation of the phase ellipse as the *twirl tw* of the dislocation line (compared to the twist Tw of the phase lines). It is found using a Stokes parameter representation of the anisotropy ellipse in its plane (the (x, y) plane), where the rotation is the rate of change of half of the Poincaré sphere azimuth $\beta = \arctan S_2/S_1$, where the Stokes parameters S_1, S_2 of the ellipse defined by the complex vector $\nabla\psi$ are defined in equation (A.4.7). The twirl is therefore

$$
\begin{aligned}
tw_\phi &= \frac{1}{2}\partial z \arctan \frac{S_2}{S_1} \\
&= \frac{S_1 S_2' - S_2 S_1'}{S_1^2 + S_2^2},
\end{aligned} \tag{2.8.9}
$$

which does not simplify particularly (although the denominator is $G^2 - 4\omega^2$). It is zero for the wave (2.8.3), since the phase ellipse is fixed with respect to the (x, y) axes. The rectifying phase χ_0, such that $\nabla\psi \exp(i\chi_0/2) = \mathbf{X}_0 + i\mathbf{Y}_0$ (A.2.6) also changes along the ellipse in general, giving the *phase twirl tw_χ*,

$$
tw_\chi = -\frac{1}{2}\partial_z \arctan \frac{2\mathbf{X} \cdot \mathbf{Y}}{X^2 - Y^2}. \tag{2.8.10}
$$

This measures the phase twist with respect to the ellipse semiaxes. Neither of these angles is defined when the ellipse is circular (see chapter 4), and neither type of twirl is defined in this nongeneric case. Another measure of the phase twist Tw_{tw} is the difference $tw_\phi - tw_\chi$, the twirl (measuring the rate of rotation of the ellipse along the dislocation line with respect to parallel transport) plus the rate at which the phase changes with respect to the ellipse. It is readily shown to be

$$
\begin{aligned}
Tw_{tw} &= -\frac{\mathbf{T} \cdot (\mathbf{X} \wedge \mathbf{X}' + \mathbf{Y} \wedge \mathbf{Y}) - (\mathbf{X} \cdot \mathbf{Y}' - \mathbf{Y} \cdot \mathbf{X}')}{X^2 + Y^2 + 2\mathbf{X} \wedge \mathbf{Y}} \\
&= -\frac{\mathrm{Re}\{\mathbf{T} \cdot (\nabla\psi^* \wedge \nabla\psi') + i\nabla\psi^* \cdot \nabla\psi'\}}{G + 2\omega}.
\end{aligned} \tag{2.8.11}
$$

Note that, although the twirls are not defined if the ellipse is circular, Tw_{tw}, is. It is appealing in other ways: it does not require averaging over χ or ϕ, and the numerator is

the difference of the numerators of tw_ϕ (2.8.7) and σ (2.8.8). The topological implications of twist and twirl are considered in chapter 5.

2.9   Discussion

This chapter sets the scene for the later chapters: values of the geometric quantities found here, such as speed and core ellipse eccentricity, are averaged in the next chapter (in an ensemble of isotropic random waves), and the topology of dislocation reconnection is used in the knot and link constructions of chapter 5. Moreover, polarization singularities in vector and tensor waves are realised as phase singularities in chapters 4, 6, and have the structures and morphologies described here. Most of the discussion in this chapter is not dependent on the field satisfying a wave equation, and applies to phase singularities in any complex field.

Although much of the work here is a review of earlier work, there is new understanding on several topics, most significantly for the topology of dislocation lines. There is still a lot to understand: can benign crossings of dislocation lines occur in the wave equation? Is the normal form (2.6.1) completely general? What geometric properties distinguish the different measures of twist in section 2.8? Can a Burgers vector be associated with a general dislocation line?

The understanding of much dislocation morphology has been aided by concrete examples of dislocations satisfying the wave equation (this was the approach of [NB74]), and a solution to many unanswered questions here would be aided by finding appropriate waves. An example would be a twisting wave dislocation with different values for the various twists of section 2.8.

Chapter 3

Dislocations in isotropic random waves

Below, a myriad, myriad waves hastening, lifting up their necks,
Waves, undulating waves, liquid, uneven, emulous waves,
Toward that whirling current, laughing and buoyant with curves

from Walt Whitman, *After the sea-ship*, in *Leaves of Grass,* 1892

It is fortuitous that the central limit theorem of mathematics (see, for example, [Fel50]) provides such a tractable probability distribution as the gaussian for many physical examples of random phenomena, and the isotropic random wave model we use here utilises gaussian random functions. The averages of many geometrical quantities associated with wave dislocations may therefore be found analytically. In this chapter the model is introduced and described, and the distributions or mean values of many of the quantities described in the previous chapter are calculated. There are three sections, not including the concluding discussion : in the first, the random wave model is defined, the second, the averages for dislocation points in two dimensions are derived, and the third, averages for dislocation lines in three dimensions. The technical details to several of the calculations are omitted here, and may be found in [BD00]. Most of the calculations and discussion here have been published in [BD00, Den01a, Den01b]. The theory of random scalar waves discussed here is generalised in the next chapter to vector waves.

3.1   The isotropic random wave model

It seems that the first investigation into the statistical geometry of random functions was that of Rice [Ric44, Ric45], who used a superposition of one dimensional sinusoidal waves with random phases and amplitudes to model the random currents in noisy circuits (such as telegraph signals), deriving expressions for densities and correlations of zeros and turning points for the functions.

Longuet-Higgins not only extended Rice's work on one-dimensional functions ([LH58, CLH56]), but also introduced the corresponding two-dimensional real random plane wave superposition [LH57b, LH57a, LH60], the complex analogue to which is used here; the phenomenon he was modelling was waves on the surface of the sea. The mathematical objects calculated in these papers, such as lengths of contour line per unit area, and densities and speeds of maxima and minima, are very close to the calculations made here, and we adopt the same methods. Dislocations were first subjected to statistical analysis by Berry [Ber78], and some of these results shall appear in the present analysis.

Gaussian random wavefields have found wide application in optics, where, as stated in chapter 1, complex wavefields are most natural, and the theory of complex gaussian random waves is particularly appropriate to the statistical analysis of laser speckle patterns [Goo75, Dai76], and there are a number of studies, experimental, numerical and theoretical (eg [BZM$^+$81, Fre98a, WH82]) of the statistical properties of dislocations and critical points in optical fields. Another physical application of random complex wavefields is in so-called quantum billiards, where the eigenfunctions of modes in cavities (in two dimensions) whose classical dynamics is chaotic are assumed to be gaussian random wave superpositions [Ber77], and are complex in the presence of in/out channels or a magnetic field [BR86, BPSS99].

Not only is the geometry of gaussian random fields important in view of its physical applications, but also from mathematical interest alone [Adl81], if only as an application of probability theory to classical differential geometry. Such fields also give a feeling for what phenomena really are generic, and provide a means of studying dislocations and critical points 'in the wild'. General accounts of the theory of calculating with gaussian random functions may be found in [Goo85, Fri95], as well as the papers cited above.

The complex fields ψ studied in this chapter are all stationary and isotropic, but otherwise quite general. Calculations are made for arbitrary power spectra (defined below), and

we shall see that the averages of geometric quantities of the field, such as the dislocation density, are determined completely by the (radial) moments of the power spectrum.

A gaussian random wavefield is taken to be a sample function of the ensemble of infinitely many scalar complex homogeneous nondispersive plane waves with speed c, in two or three dimensions,

$$\psi(\mathbf{r},t) \;=\; \sum_{\mathbf{k}} a_{\mathbf{k}} \exp(\mathrm{i}(\mathbf{k}\cdot\mathbf{r} - ckt)), \qquad (3.1.1)$$

$$\psi(\mathbf{R},t) \;=\; \sum_{\mathbf{K}} a_{\mathbf{K}} \exp(\mathrm{i}(\mathbf{K}\cdot\mathbf{R} - cKt)), \qquad (3.1.2)$$

where the sum is over all wavevectors $\mathbf{k} = (k_x, k_y, k_z)$ in three dimensions and $\mathbf{K} = (K_x, K_y)$ in two (conventions are described in section 1.9). Reference shall usually be made to the three dimensional waves (3.1.1), but the two-dimensional analogue ought to be obvious. The complex amplitudes multiplying the plane waves are, in (3.1.1),

$$a_{\mathbf{k}} = \varepsilon_k \exp(\mathrm{i}\phi_{\mathbf{k}}), \qquad (3.1.3)$$

and each $\phi_{\mathbf{k}}$ (labelled by the wavevector $\mathbf{k}$) is uniformly distributed in the interval $0 \le \phi_{\mathbf{k}} < 2\pi$. The nonnegative real amplitude ε_k can have a Rayleigh distribution, the calculations here are unaffected whether it does or not. ε_k is only dependent on the wavenumber $k = |\mathbf{k}|$, so the functions (3.1.1), (3.1.2) are statistically isotropic (rotation invariant), and all averages are independent of any particular direction.

The ensembles (3.1.1), (3.1.2) are also manifestly stationary (statistically translation invariant) both in space and time, since any translation is absorbed by the random phase $\phi_{\mathbf{k}}$. They are also ergodic, so spatial (or temporal) averages commute with ensemble averages. Averaging over the ensemble of $\phi_{\mathbf{k}}$ shall be denoted by angle brackets, so formally, for any function f,

$$\langle f \rangle = \left(\prod_{\mathbf{k}} \frac{1}{2\pi} \int \mathrm{d}\phi_{\mathbf{k}} \right) f. \qquad (3.1.4)$$

Due to the uniformly random phases $\phi_{\mathbf{k}}$, it follows that the real and imaginary parts

$$\xi(\mathbf{r},t) \;=\; \sum_{\mathbf{k}} \varepsilon_k \cos(\mathbf{k}\cdot\mathbf{r} - ckt + \phi_{\mathbf{k}}) \qquad (3.1.5)$$

$$\eta(\mathbf{r},t) \;=\; \sum_{\mathbf{k}} \varepsilon_k \sin(\mathbf{k}\cdot\mathbf{r} - ckt + \phi_{\mathbf{k}}) \qquad (3.1.6)$$

are identically gaussian distributed, so ψ satisfies circular gaussian statistics [Goo85]. The statistics for ψ are also phase-stationary, and are unaffected by a global change in the

phase (gauge transformation). The variance of ξ (and identically for η) is

$$
\begin{aligned}
\langle \xi^2 \rangle &= \sum_{\mathbf{k}} \sum_{\mathbf{k}'} \varepsilon_k \varepsilon_{k'} \langle \cos(\mathbf{k} \cdot \mathbf{r} - ckt + \phi_{\mathbf{k}}) \cos(\mathbf{k}' \cdot \mathbf{r} - ck't + \phi_{\mathbf{k}'}) \rangle \\
&= \sum_{\mathbf{k}} \varepsilon_k^2 \langle \cos^2(\mathbf{k} \cdot \mathbf{r} - ckt + \phi_{\mathbf{k}}) \rangle \\
&= \frac{1}{2} \sum_{\mathbf{k}} \varepsilon_k^2,
\end{aligned} \tag{3.1.7}
$$

and assume that the sum in $\mathbf{k}$ is sufficiently finely spaced that it may be replaced by an integral,

$$
\begin{aligned}
\frac{1}{2} \sum_{\mathbf{k}} \varepsilon_k^2 &\equiv \int \mathrm{d}^3 \mathbf{k} E(k) \\
&= \int \mathrm{d}k \frac{\Pi(k)}{4\pi k^2}
\end{aligned} \tag{3.1.8}
$$

where E is the *energy spectrum* of the wave, and Π is the radial energy spectrum (we use the term 'energy spectrum' rather than 'power spectrum', which implies a time dependence we shall not usually be interested in). E and Π only depend on wavenumber k (since ε_k does), so Π is the natural distribution to consider, and it has a finite integral (otherwise the variance (3.1.7) would diverge). Averages of k-dependent quantities with respect to the power spectrum are denoted by double angle brackets $\langle\langle \bullet \rangle\rangle$, and the nth moment of k is written k_n, that is

$$
k_n = \langle\langle k^n \rangle\rangle = \int_0^\infty \mathrm{d}k \, k^n \Pi(k), \tag{3.1.9}
$$

and $\Pi(k)$ is taken to be normalised, so $k_0 = 1$. Rescaling Π (equivalently, ψ) does not alter any of the physical averages here. For two-dimensional wavefields, the two dimensional radial energy spectrum $\Pi_2(K)$ is defined

$$
\frac{1}{2} \sum_{\mathbf{K}} \varepsilon_K^2 \equiv \int \mathrm{d}^2 \mathbf{K} \frac{\Pi_2(K)}{2\pi K}, \tag{3.1.10}
$$

and where confusion might ensue, subscripts 2,3 shall be used appropriately on the radial spectrum Π and its averages $\langle\langle \bullet \rangle\rangle$. The moments K_n are defined analogously to those of equation (3.1.9), and Π_2 is normalised such that $K_0 = 1$.

If ψ is a plane section of an isotropic wavefield in three dimensions, the Π_3, Π_2 are related by projection in wavevector space (where the plane is taken to be the (k_x, k_y)

plane),

$$
\begin{aligned}
\Pi_2(K) &= 2\pi K \int \mathrm{d}k_z \frac{\Pi_3(\sqrt{K^2 + k_z^2})}{4\pi(k_z^2 + K^2)} \\
&= K \int_K^\infty \mathrm{d}k \frac{\Pi_3(k)}{k\sqrt{k^2 - K^2}},
\end{aligned}
\tag{3.1.11}
$$

and it is easily shown that for such a projection

$$
K_n = \frac{(\frac{1}{2})!(\frac{n}{2})!}{(\frac{n+1}{2})!} k_n.
\tag{3.1.12}
$$

$P(f)$ is understood to represent the probability density function for a quantity f with respect to the ensemble of functions ψ in (3.1.1), (3.1.2). The normalisation of Π, and the circular gaussian statistics for ψ, imply that

$$
\begin{aligned}
P(\psi) &= \frac{1}{2\pi} \exp(-|\psi|^2/2) \tag{3.1.13} \\
P(\xi, \eta) &= \frac{1}{2\pi} \exp(-(\xi^2 + \eta^2)/2), \tag{3.1.14}
\end{aligned}
$$

and probabilities usually will only be taken for real functions (that is, like (3.1.14) rather than (3.1.13)).

In addition to averaging geometric quantities, correlation functions of dislocation positions shall also be found for dislocation points in the plane, where values of ψ are taken at two points $\mathbf{R}_A$ and $\mathbf{R}_B \equiv \mathbf{R}_A + \mathbf{R}$; values of functions at these two points are written with subscript A or B appropriately (eg $\xi(\mathbf{R}_A) = \xi_A$), and by isotropy, it is only the magnitude $R = |\mathbf{R}|$ that is important. The simplest such average is the (spatial) *autocorrelation function $C(R)$*, defined

$$
C(R) = \langle \xi_A \xi_B \rangle = \langle \eta_A \eta_B \rangle,
\tag{3.1.15}
$$

and by an argument similar to (3.1.7), by (3.1.8), (3.1.10),

$$
C(R) = \langle\langle J_0(KR)\rangle\rangle_2 = \left\langle\left\langle \frac{\sin kR}{kR} \right\rangle\right\rangle_3,
\tag{3.1.16}
$$

where J_0 is the zero order Bessel function of the first kind. Note that the autocorrelation function, as defined here, is real, since it is the autocorrelation function of the real functions ξ, η; other authors define the autocorrelation function as $\langle\langle \psi_A \psi_B \rangle\rangle$. The Wiener-Khinchin theorem [Goo85, Fri95] states that the autocorrelation function is the two or three-dimensional inverse Fourier transform of the energy spectrum E, implying that the moments k_n, K_n are related to the derivatives of $C(R)$ at the origin. The normalisation implies that $C(0) = 1$.

The moments k_n, K_n shall appear in calculations as quadratic averages of ξ, η and their (spatial and time) derivatives which are also stationary gaussian random functions. The only nonvanishing averages (up to second order in space, first in time) are, for three dimensions, where $i \neq j$ denote x, y, z,

$$\langle \xi_i^2 \rangle = \langle \eta_i^2 \rangle = -\langle \xi \xi_{ii} \rangle = -\langle \eta \eta_{ii} \rangle = k_2/3$$

$$\langle \xi_{ii}^2 \rangle = \langle \eta_{ii}^2 \rangle = k_4/5$$

$$\langle \xi_{ij}^2 \rangle = \langle \eta_{ij}^2 \rangle = \langle \xi_{ii} \xi_{jj} \rangle = \langle \eta_{ii} \eta_{jj} \rangle = k_4/15$$

$$\langle \xi_t^2 \rangle = \langle \eta_t^2 \rangle = c^2 k_2$$

$$\langle \xi \eta_t \rangle = -\langle \eta \xi_t \rangle = c k_1, \tag{3.1.17}$$

and in two dimensions, where $\alpha \neq \beta$ denote x, y,

$$\langle \xi_\alpha^2 \rangle = \langle \eta_\alpha^2 \rangle = -\langle \xi \xi_{\alpha\alpha} \rangle = -\langle \eta \eta_{\alpha\alpha} \rangle = K_2/2$$

$$\langle \xi_{\alpha\alpha}^2 \rangle = \langle \eta_{\alpha\alpha}^2 \rangle = 3K_4/8$$

$$\langle \xi_{\alpha\beta}^2 \rangle = \langle \eta_{\alpha\beta}^2 \rangle = \langle \xi_{\alpha\alpha} \xi_{\beta\beta} \rangle = \langle \eta_{\alpha\alpha} \eta_{\beta\beta} \rangle = K_4/8$$

$$\langle \xi_t^2 \rangle = \langle \eta_t^2 \rangle = c^2 K_2 \qquad \text{(or, for plane sections of spatial waves, } c^2 k_2)$$

$$\langle \xi \eta_t \rangle = -\langle \eta \xi_t \rangle = c K_1 \qquad \text{(or, for plane sections of spatial waves, } c k_2), \tag{3.1.18}$$

which are consistent with (3.1.12). Note that, if the two-dimensional wavefield is a section of a spatial wave, the time derivative correlations involve the spatial k (because they come from an angular frequency).

For averages involving the correlation function C, it matters in what direction spatial derivatives are taken with respect to the vector $\mathbf{R}$. For convenience, we take $\mathbf{R}$ to be in the x-direction (and isotropy is recovered after averaging). Therefore, where $\bullet' = \mathrm{d}/\mathrm{d}R$,

$$\langle \xi_A \xi_{Bx} \rangle = \langle \eta_A \eta_{Bx} \rangle = -\langle \xi_{Ax} \xi_B \rangle = -\langle \eta_{Ax} \eta_B \rangle = -\langle\langle K J_1(KR) \rangle\rangle_2 = C'(R)$$

$$\langle \xi_{Ax} \xi_{Bx} \rangle = \langle \eta_{Ax} \eta_{Bx} \rangle = -\langle\langle K^2 J_0''(KR) \rangle\rangle_2 = -C''(R)$$

$$\langle \xi_{Ay} \xi_{By} \rangle = \langle \eta_{Ay} \eta_{By} \rangle = -\langle\langle K^2 J_1(KR) \rangle\rangle_2/R = -C'(R)/R, \tag{3.1.19}$$

where the three-dimensional autocorrelation is easy to find, and noting the small correction from [BD00] equation (3.16).

All other averages (up to appropriate orders) vanish; in particular, the only nonvanishing average involving ξ, η together is $\langle \xi \eta_t \rangle = -\langle \eta \xi_t \rangle$.

These correlations are used to average over the various geometric quantities for dislocations, found in the previous chapter as functions of $\xi, \eta, \nabla\xi, \nabla\eta$, etc. The theory of integrating gaussian random functions is well known (see, for example, [Goo85, Fri95]), and the techniques are only outlined here. Consider any vectoral set of N gaussian random scalar functions, depending on space, time, and the choice of random phases ϕ_k,

$$\mathbf{u}(\mathbf{r}, t) = (u_1(\mathbf{r}, t), u_2(\mathbf{r}, t), \ldots, u_N(\mathbf{r}, t)), \tag{3.1.20}$$

(which may be ξ, η, ξ_x, η_y, etc), and a vectoral set of auxiliary variables

$$\mathbf{b} = (b_1, b_2, \ldots, b_n). \tag{3.1.21}$$

Then, using gaussian randomness, it can be shown that

$$\langle \exp(\mathrm{i}\mathbf{b} \cdot \mathbf{u}(\mathbf{r}, t)) \rangle = \exp(-\frac{1}{2}\langle (\mathbf{b} \cdot \mathbf{u})^2 \rangle) = \exp(-\frac{1}{2}\mathbf{b} \cdot \mathbf{M} \cdot \mathbf{b}), \tag{3.1.22}$$

where $\mathbf{M}$ is the *correlation matrix* with elements

$$M_{mn} \equiv \langle u_m u_n \rangle, \tag{3.1.23}$$

found in the appropriate tables above. Using (3.1.22), it is possible to find the joint probability density function $P(u_1, \ldots, u_N)$ of the value of the gaussian random variables u_n, by taking the Fourier transform of the δ-function:

$$\begin{aligned} P(\mathbf{u}) &\equiv \langle \delta(\mathbf{u} - \mathbf{u}(\mathbf{r}, t)) \rangle \\ &= \frac{1}{2\pi} \int \mathrm{d}^N \mathbf{b} \, \exp(-\mathrm{i}\mathbf{b} \cdot \mathbf{u}) \langle \exp(\mathrm{i}\mathbf{b} \cdot \mathbf{u}(\mathbf{r}, t)) \rangle \\ &= \frac{\exp(-\mathbf{u} \cdot \mathbf{M}^{-1} \cdot \mathbf{u})}{(2\pi)^{N/2}\sqrt{\det \mathbf{M}}} \end{aligned} \tag{3.1.24}$$

The use of these quadratic averages of the field is used to calculate the joint probability for the quantities $G = (\nabla\xi)^2 + (\nabla\eta)^2$, $\omega = |\nabla\xi \wedge \nabla\eta|$ from equations (2.2.2), (2.3.5), defined appropriately in two or three dimensions. The probability distributions are

$$P_3(G, \omega) = \frac{27\omega}{2k_2^3} \exp(-3G/2k_2)\Theta(G - 2\omega), \tag{3.1.25}$$

$$P_2(G, \omega) = \frac{1}{K_2^2} \exp(-G/K_2)\Theta(G - 2|\omega|), \tag{3.1.26}$$

and details of these derivations may be found in [BD00] appendix A. Recall that $2\omega/G$ is the normalised Stokes parameter s_3 for the ellipse described by $\nabla\psi$ (equations (A.4.7),

(A.4.8)); the density (3.1.26) shows that this Stokes parameter is uniformly distributed between $(-1, 1)$, regardless of the spectrum (Stokes parameter statistics are considered in more detail in the next chapter).

The remainder of this section is a description of five particular wave spectra of theoretical or experimental interest, whose dislocation statistics shall be calculated explicitly. Three are for waves in the plane, two for waves in space (whose planar statistics are also evaluated):

(1) *Monochromatic waves in the plane.* The energy spectrum is a δ-function at some fixed wavenumber K_{m}, wavelength $\Lambda_{\mathrm{m}} = 2\pi/K_{\mathrm{m}}$ and the field satisfies the planar Helmholtz equation (1.5.2), (2.2.15). Following [LH57a], we call the spectrum for this ensemble the *ring spectrum*, since all of the waves in Fourier space lie on a ring of radius K_{m}. Thus

$$\Pi_{\mathrm{ring}}(K) = \delta(K - K_{\mathrm{m}}), \tag{3.1.27}$$

implying that

$$K_{n\ \mathrm{ring}} = K_{\mathrm{m}}^{n} \tag{3.1.28}$$

and

$$C_{\mathrm{ring}}(R) = J_0(K_{\mathrm{m}}R), \tag{3.1.29}$$

easily identifiable as the two-dimensional Fourier transform of a ring. These waves are the most coherent planar waves possible since there is only one wavelength present, and the Bessel function gives interesting long range properties, as discussed in [OGH87, HOG89]. [Ber77] suggested that these wavefields are a good model for quantum billiards, since the ray dynamics for such systems, corresponding to the directions of the wavevectors, is random. Dislocations in such quantum waves have been studied by [BR86, MB89, BPSS99, SBS01]. The radial spectrum, autocorrelation function, and intensity and real, imaginary zero contours of a random sample function (a sum of 100 monochromatic plane waves with random directions and phases) for the ring spectrum are shown in figure (3.1).

(2) *Disk spectrum.* In this case the energy spectrum is a radial step function of radius $K_{\mathrm{d}} = 2\pi/\Lambda_{\mathrm{d}}$, so has radial energy spectrum

$$\Pi_{\mathrm{disk}}(K) = 2K\Theta(K - K_{\mathrm{d}})/K_{\mathrm{d}}^2. \tag{3.1.30}$$
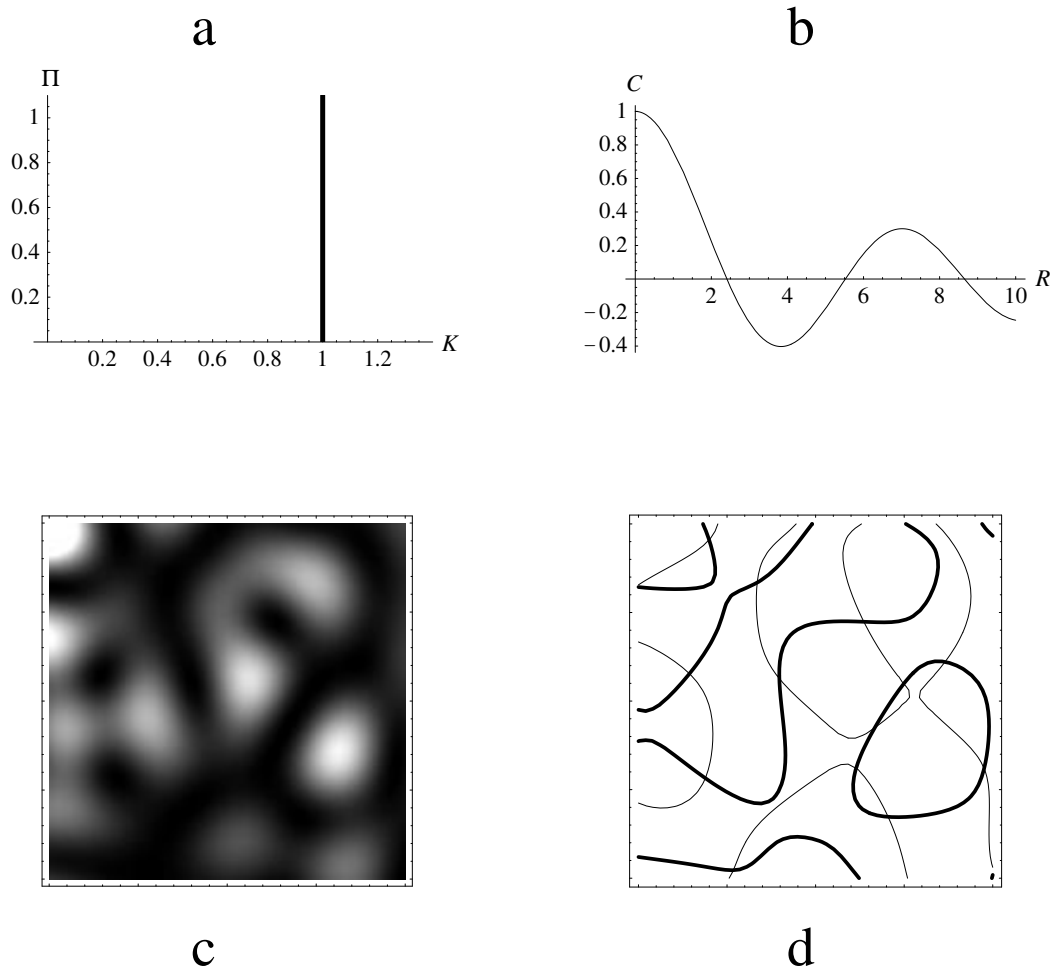
Figure 3.1: The ring spectrum: (a) Radial spectral density (δ-function), (units of K_m), (b) Autocorrelation function J_0 (in units of $1/K_\mathrm{m}$), (c) Intensity (light is more intense) of a sample function, (d) $\xi = 0, \eta = 0$ contours for the same sample function, plot size is four square wavelengths.

This is a spectrum often studied in optics, being that of uniform diffuse monochromatic light in three dimensions illuminating a plane, having passed through a circular aperture (the geometry of the energy spectrum matching the aperture shape by the Zernike-van Cittert theorem [Goo85, BW59]). Note that although the two-dimensional $\Pi_{\mathrm{disk}}(K)$ is not monochromatic, it arises from the interference of a spatial monochromatic wave, so there is only one angular frequency component ω_k, and the dislocation pattern is therefore stationary, as discussed in section 1.5. The disk spectrum has moments

$$K_{n\ \mathrm{disk}} = 2K_{\mathrm{d}}^n/(2+n), \tag{3.1.31}$$

and autocorrelation function

$$C_{\mathrm{disk}}(R) = 2J_1(K_{\mathrm{d}}R)/K_{\mathrm{d}}R, \tag{3.1.32}$$

the inverse Fourier Transform of a disk in K-space. Dislocations with the disk spectrum have been investigated numerically by Freund [Fre98a]. The spectrum, autocorrelation function and a sample function of the disk spectrum are shown in figure (3.2).

(3) *Gaussian spectrum.* The energy spectrum is a gaussian, with standard deviation $K_\sigma = 2\pi/\Lambda_\sigma$, so

$$\Pi_{\mathrm{gauss}}(K) = K\exp(-K^2/2K_\sigma^2)/K_\sigma^2. \tag{3.1.33}$$

This is a reasonable model for the monochromatic speckle pattern transverse to a paraxial laser beam, or from a gaussian scatterer. As with the disk spectrum, the physical (spatial) beam is monochromatic, so the dislocation pattern is stationary. This distribution has moments

$$K_{n\ \mathrm{gauss}} = 2^{n/2}\left(\frac{n}{2}\right)!K_\sigma^n \tag{3.1.34}$$

and gaussian autocorrelation function

$$C_{\mathrm{gauss}}(R) = \exp(-K_\sigma^2R^2/2), \tag{3.1.35}$$

and is shown in figure (3.3). Some statistical features of these fields were investigated by [OG83].
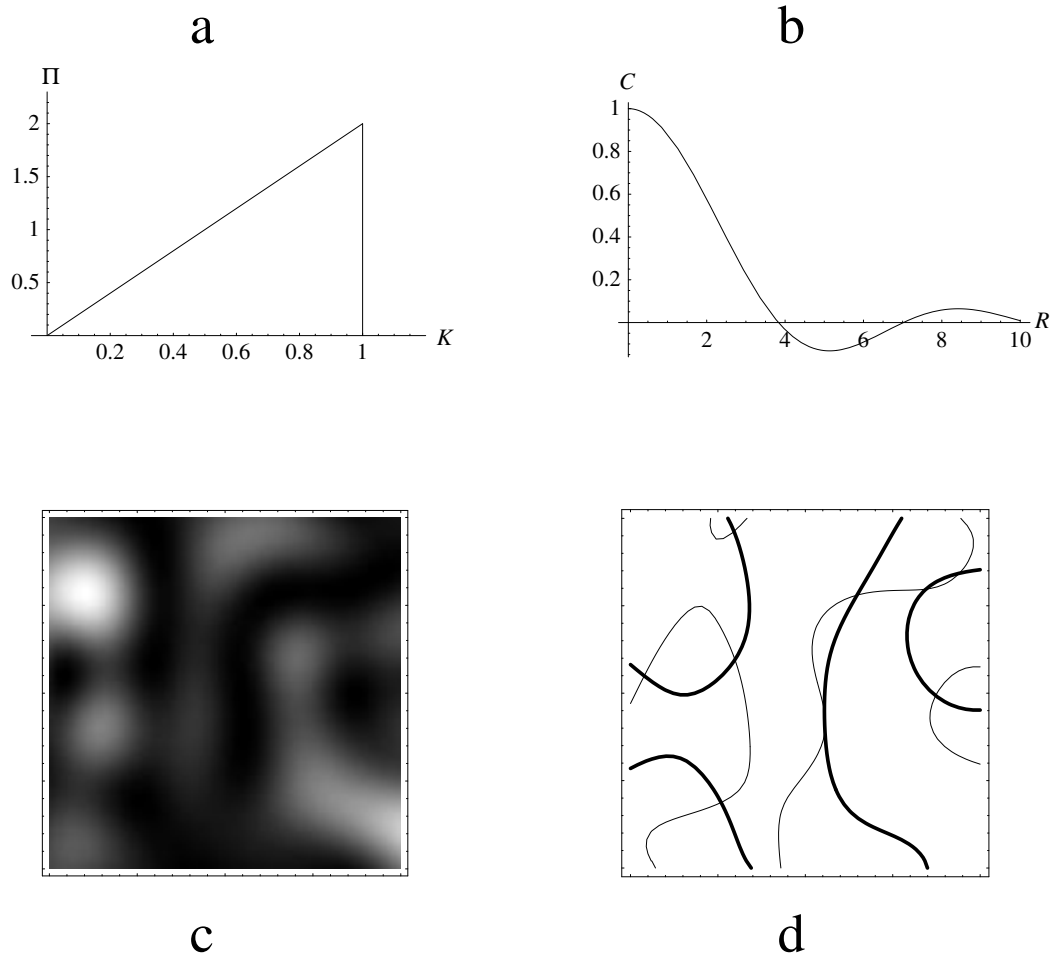
Figure 3.2: The disk spectrum: (a) Radial spectral density, in units of K_{d}, (b) Autocorrelation function, in units of $1/K_{\mathrm{d}}$, (c) Intensity (light is more intense) of a sample function, (d) $\xi = 0, \eta = 0$ contours for the same sample function (plot size is four square wavelengths Λ_{d}.
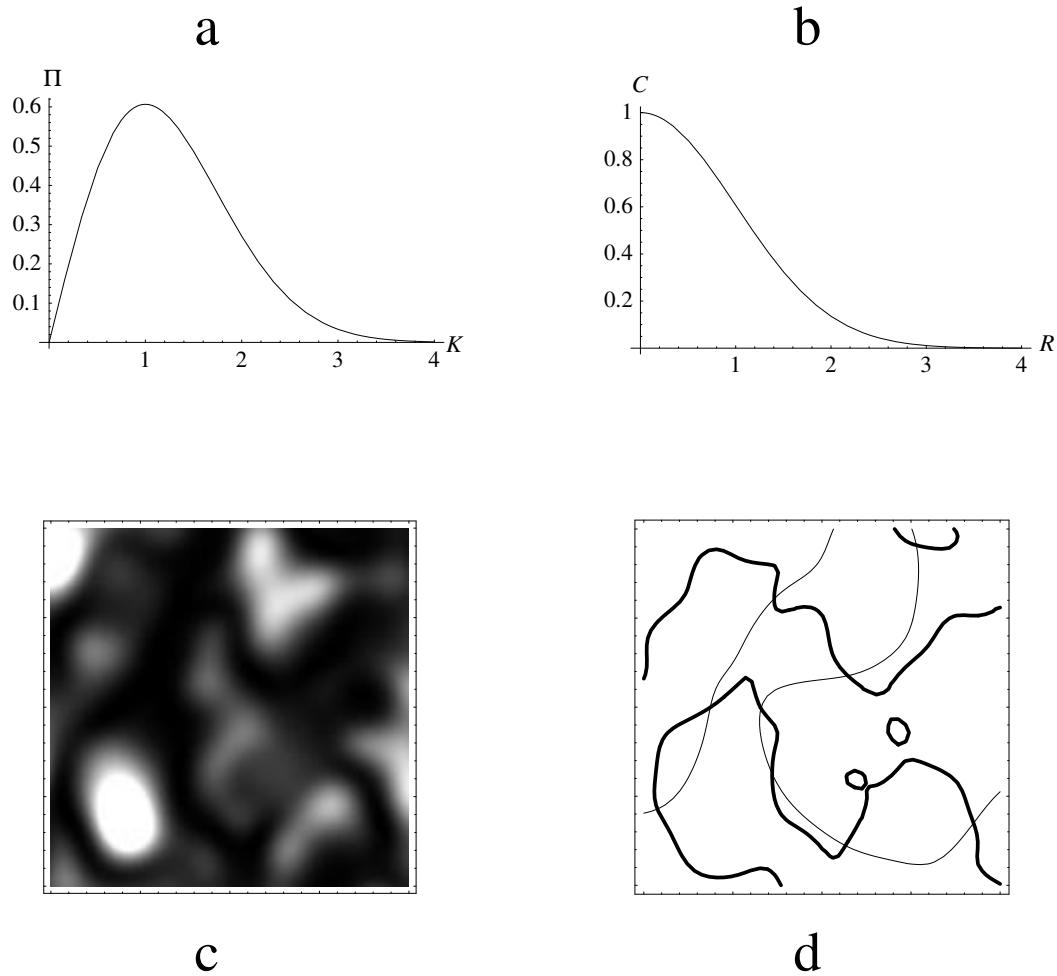
Figure 3.3: The gaussian spectrum: (a) Radial spectral density (units of K_σ), (b) Autocorrelation function (units of $1/K_\sigma$), (c) Intensity (light is more intense) of a sample function, (d) $\xi = 0, \eta = 0$ contours for the same sample function, in four square wavelengths.

(4) *Monochromatic waves in space.* This is the three-dimensional analogue of the ring spectrum, where all of the waves in the superposition (3.1.1) have wavenumber $k_{\mathrm{m}} = 2\pi/\lambda_{\mathrm{m}}$. It shall be referred to as the *shell spectrum*, and has the three-dimensional radial spectrum

$$\Pi_{\mathrm{shell}}(k) = \delta(k - k_{\mathrm{m}}), \tag{3.1.36}$$

and spatial moments

$$k_{n\ \mathrm{shell}} = k_{\mathrm{m}}^n. \tag{3.1.37}$$

By equation (3.1.11), the shell spectrum projects to a radial planar spectrum

$$\Pi_{2\ \mathrm{shell}}(K) = \frac{K\Theta(k_{\mathrm{m}} - K)}{k_{\mathrm{m}}\sqrt{k_{\mathrm{m}}^2 - K^2}} \tag{3.1.38}$$

and planar moments found directly from equation (3.1.12). The autocorrelation function is the spherical Bessel function

$$C_{\mathrm{shell}} = \frac{\sin k_{\mathrm{m}}R}{k_{\mathrm{m}}R}. \tag{3.1.39}$$

Such a wavefield, shown in figure (3.4), is a model for monochromatic waves in a three-dimensional chaotic cavity.

(5) *The Planck spectrum.* This is the Planck's distribution for blackbody radiation, which has the normalised radial spectrum

$$\Pi_{\mathrm{Planck}}(k) = \frac{15k^3}{\pi^4 k_T^4(\exp(k/k_T) - 1)}, \tag{3.1.40}$$

where the thermal wavelength k_T and thermal wavenumber λ_T are defined, for temperature T,

$$k_T = \frac{k_{\mathrm{B}}T}{\hbar c}, \qquad \lambda_T = \frac{hc}{k_{\mathrm{B}}T}, \tag{3.1.41}$$

where k_{B} is Boltzmann's constant, $h = 2\pi\hbar$ is Planck's constant and c is the speed of light in vacuo. It is easily shown that the moments of this spectrum are

$$k_{n\ \mathrm{Planck}} = 15(n + 3)!\zeta(n + 4)/\pi^4, \tag{3.1.42}$$

where ζ here represents the Riemann ζ-function. The corresponding planar moments $K_{n\mathrm{Planck}}$ are easy to determine by equation (3.1.12), although there is no simple analytic form for the plane projected spectrum $\Pi_{2\mathrm{Planck}}$. The autocorrelation function, correctly calculated in the errata to [BD00], is

$$C_{\mathrm{Planck}}(R) = \frac{15}{(\pi k_T R)^4}\left(1 - \frac{(\pi k_T R)^3 \cosh(\pi k_T R)}{\sinh^3(\pi k_T R)}\right). \tag{3.1.43}$$
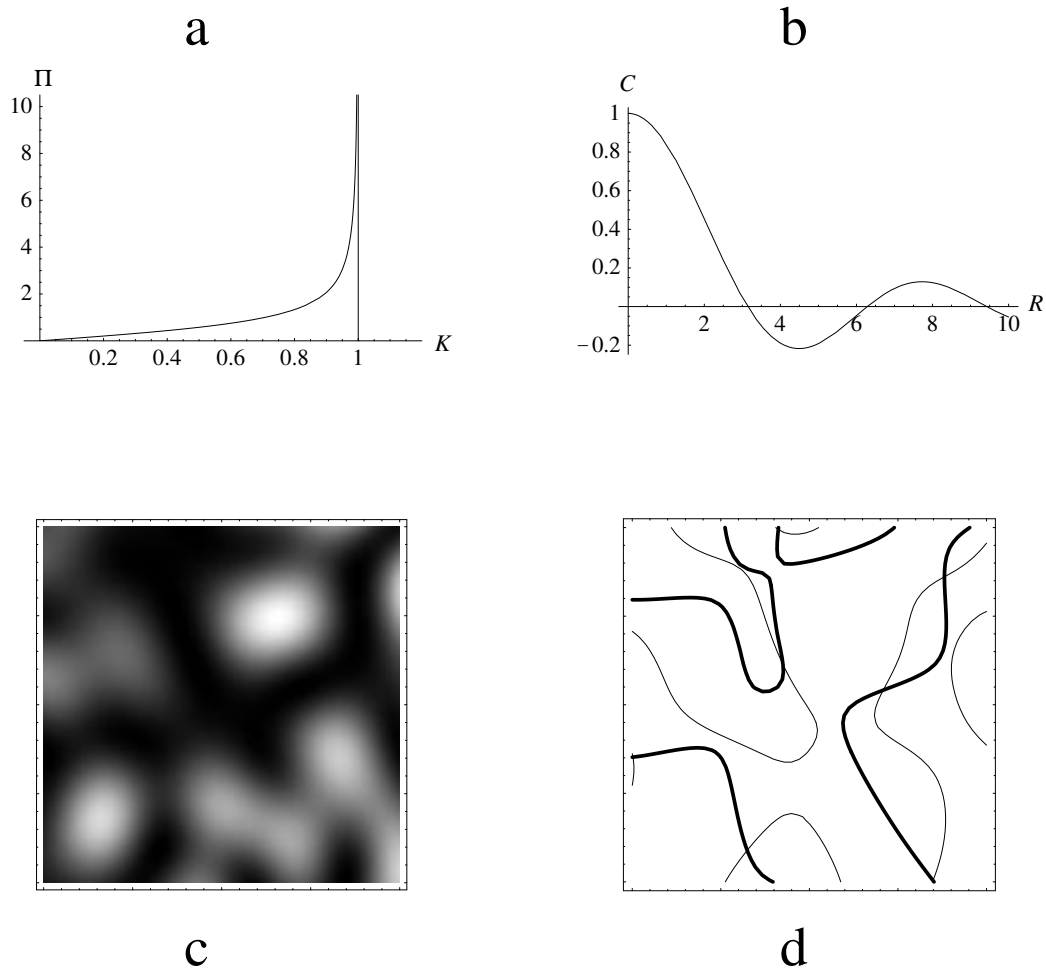
Figure 3.4: The shell spectrum: (a) Plane projected radial spectral density $\Pi_{2\text{shell}}(K)$ (units k_m), (b) Autocorrelation function (units $1/k_\text{m}$), (c) Intensity (light is more intense) of a plane section of a sample function, (d) $\xi = 0, \eta = 0$ contours for the same plane section, for four square wavelengths.
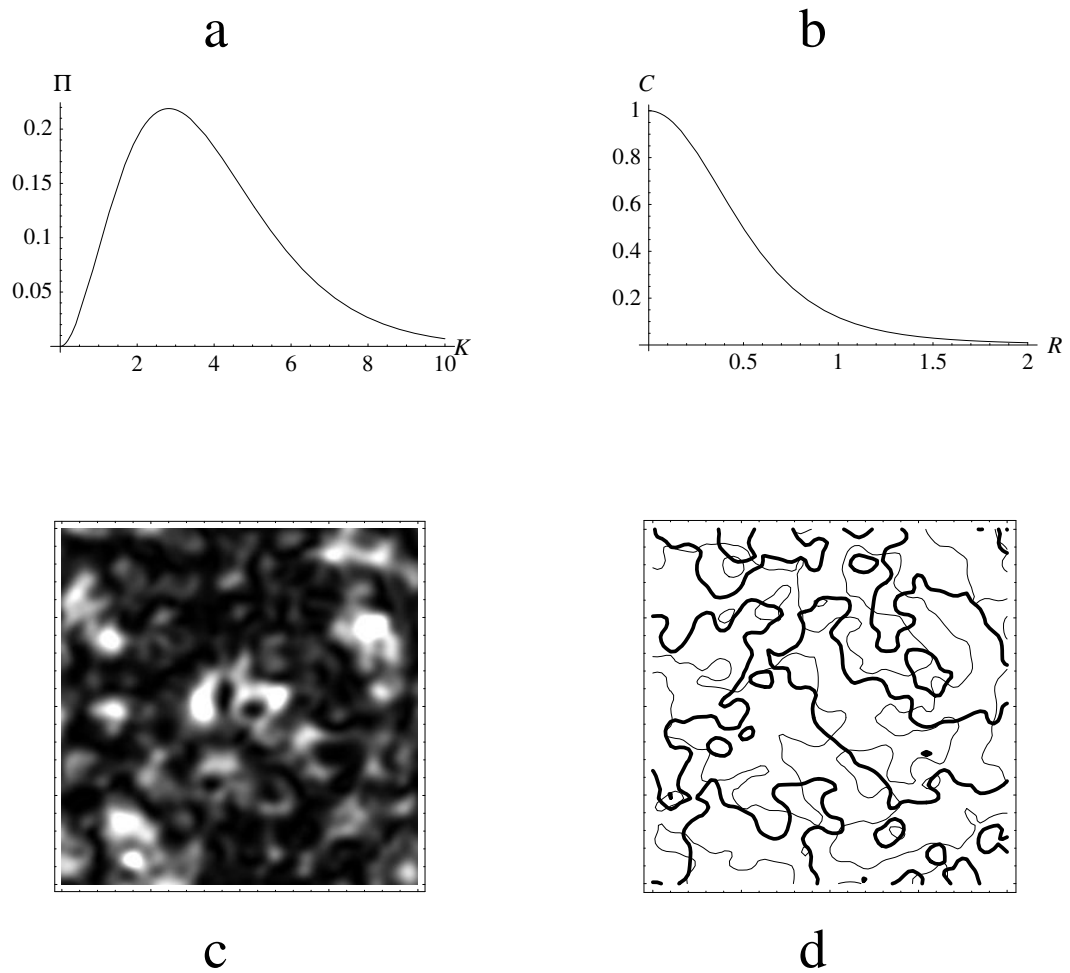
Figure 3.5: The Planck spectrum: (a) Radial spectral density (not projected, in units of k_T; note the peak at 2.82), (b) Autocorrelation function (units $1/k_T$), (c) Intensity of a plane section of a sample function, (d) $\xi = 0, \eta = 0$ contours for the same plane section, in four square thermal wavelengths.

The scalar caricature of blackbody radiation has historically been used several times (such as [Ray89, EH10a, EH10b]), and certainly is reasonable as a model (in chapter 4 we shall consider blackbody vector radiation). Bourret [Bou60] studied the longitudinal and lateral autocorrelation functions of vector waves with the Planck spectrum, finding functions similar (but not identical) to equation (3.1.43). Note that, as in figure (3.5), the peak of Π_{Planck}, at $2.82 k_T$, is large compared to k_T, unlike the other radial spectra, whose peaks are all approximately 1 in the appropriate wavenumber units. This implies that, although k_T, λ_T are convenient units to use, the structural features (such as dislocation curvature) will all appear to be very small in these units (as is obvious, for example, in figure (3.5c)). Note that this is the only spectrum out of the five listed here whose dislocation pattern is moving, the real waves being complexified using the analytic signal of section 1.5.

3.2 Statistical geometry of dislocation points in two dimensions

This section consists of calculations of dislocation density, critical point density, dislocation correlation functions, anisotropy ellipse eccentricity and speed probability density functions, for the ensemble of isotropic gaussian random waves in the plane, defined in equation (3.1.2).

3.2.1 Dislocation density

The first quantity to be calculated is the mean density of dislocation points in two dimensions. It is the fundamental statistic for dislocations in the plane, giving an approximation of the mean spacing of dislocations (as $1/\sqrt{\text{density}}$), and is also the simplest of the calculations presented here. The number of dislocations $D_{\mathcal{A}}$ in an area $\mathcal{A}$ was given in equation (2.1.6). The average number d_D of dislocations in the plane, therefore, is the average over space (where the symbol $\mathcal{A}$ also represents the area of the set $\mathcal{A}$), taking the limit as $\mathcal{A}$ fills the space,

$$
\begin{aligned}
d_D &= \lim_{\mathcal{A} \to \mathbb{R}^2} \frac{D_{\mathcal{A}}}{\mathcal{A}} \\
&= \langle \delta(\xi)\delta(\eta)|\omega| \rangle,
\end{aligned}
\tag{3.2.1}
$$

where in the second line, the spatial average $\lim_{\mathcal{A} \to \mathbb{R}^2} 1/\mathcal{A}$ has been replaced, by ergodicity, with the ensemble average, and we recall from equation (2.2.2) that $\omega = \xi_x \eta_y - \xi_y \eta_x$. There are therefore six gaussian random functions, $\xi, \eta, \xi_x, \xi_y, \eta_x, \eta_y$ that are involved in the evaluation, which are all independent from (3.1.18). Using (3.1.24) to find the necessary probability density functions, we find

$$
\begin{aligned}
d_D &= \int \mathrm{d}\xi \mathrm{d}\eta \mathrm{d}\xi_x \mathrm{d}\xi_y \mathrm{d}\eta_x \mathrm{d}\eta_y \; \delta(\xi)\delta(\eta) |\xi_x \eta_y - \xi_y \eta_x| P(\xi, \eta, \xi_x, \xi_y, \eta_x, \eta_y) \\
&= \frac{1}{2\pi} \frac{1}{(\pi K_2)^2} \int \mathrm{d}\xi_x \mathrm{d}\xi_y \mathrm{d}\eta_x \mathrm{d}\eta_y \; |\xi_x \eta_y - \xi_y \eta_x| \exp(-(\xi_x^2 + \xi_y^2 + \eta_x^2 + \eta_y^2)/K_2) \\
&= \frac{1}{\pi^3 K_2^2} \, 2\pi \int_0^\infty \mathrm{d}X \int_0^\infty \mathrm{d}Y \int_0^{2\pi} \mathrm{d}\theta \, X^2 Y^2 |\sin\theta| \exp(-(X^2 + Y^2)/K_2) \\
&= \frac{2K_2}{\pi^2} \int_0^\pi \mathrm{d}\theta \, \sin\theta \left[\int_0^\infty \mathrm{d}X \, X^2 \exp(-X^2) \right]^2 \\
&= \frac{K_2}{4\pi}.
\end{aligned}
\tag{3.2.2}
$$

In the third line, $\nabla\xi$ and $\nabla\eta$ are transformed to polar coordinates $\nabla\xi \to (X, \theta_0), \nabla\eta \to (Y, \theta_0 + \theta)$, and ω becomes $XY\sin\theta$ under this transformation. θ_0, not appearing in any of the quantities being averaged, is integrated automatically. This technique could be used because of isotropy, and shall be frequently taken advantage of in the calculations to follow. Only the second moment appears in the result (3.2.2), having the correct dimension of inverse length squared, and is multiplied by a trigonometric factor of $1/4\pi$. Most dislocation statistics will have this form (a moment of K, or product of moments, times a trigonometric factor). The dislocation density may also be written in terms of the derivatives of the field autocorrelation function $C(R)$ when $R = 0$,

$$
d_D = \frac{K_2}{4\pi} = -\frac{C''(0)}{2\pi}.
\tag{3.2.3}
$$

It is also straightforward to calculate the mean dislocation strength density d_s by averaging the expression (2.1.7) for dislocation strength in an area,

$$
\begin{aligned}
d_s &= \langle \delta(\xi)\delta(\eta)\omega \rangle \\
&= \frac{K_2}{\pi^2} \int_0^{2\pi} \mathrm{d}\theta \, \sin\theta \times \left[\int_0^\infty \mathrm{d}X \, X^2 \exp(-X^2) \right]^2 \\
&= 0.
\end{aligned}
\tag{3.2.4}
$$

As one would expect, there is no statistical preference for either $+1$ or -1 dislocations, and the net topological charge is zero. Therefore, in addition to the topological neutrality

condition of the sign rule, there is a second, statistical neutrality condition, that of global (statistical) neutrality.

It appears that this calculation was first made by Berry [Ber78], (for slightly more general waves, including anisotropy), and was rederived by Halperin [Hal81], in a form closer to that here. The densities for the various spectra are shown in table 3.1.

Dislocation averages, that is, averages of any quantity f for dislocations in the plane (such as anisotropy ellipse eccentricity), are defined

$$\langle f \rangle_{\mathrm{d}} \equiv \frac{1}{d_D} \langle \delta(\xi)\delta(\eta)|\omega|f \rangle, \qquad (3.2.5)$$

which gives the correct statistical weighting.

3.2.2   Phase critical point density

It is also possible to evaluate the density of phase critical points (points where $\nabla\chi = 0$), described and discussed in section 2.2. The dislocations and critical points are all realised as zeros of the current $\mathbf{J}$, defined in equation (2.2.1), and, using ergodicity and equations (2.2.7), (2.2.9), the mean density d_Z of current zeros is

$$
\begin{aligned}
d_Z &= \langle \delta^2(\mathbf{J})|\det \mathbf{M_J}| \rangle \\
&= \langle \delta^2(\mathbf{J})|\det \mathbf{M_{J}}_{\mathrm{sym}}| \rangle + \langle \delta^2(\mathbf{J})|\det \mathbf{M_{J}}_{\mathrm{asym}}| \rangle \qquad (3.2.6)
\end{aligned}
$$

where $\mathbf{M_J} = \partial_\alpha J_\beta$, $\alpha, \beta = x, y$, with determinants of symmetric and antisymmetric parts given in equations (2.2.10), (2.2.11). The $\mathbf{M_{J}}_{\mathrm{asym}}$ summand in the second line of (3.2.6) merely gives the dislocation density (3.2.2), as may be verified by direct substitution of variables. The critical point density d_C (sum of saddle density d_S and extremum density d_E) is therefore

$$d_C = \langle \delta^2(\mathbf{J})|\det \mathbf{M_{J}}_{\mathrm{sym}}| \rangle. \qquad (3.2.7)$$

The following Fourier identities are used, with μ real:

$$\delta(\mu) = \frac{1}{2\pi} \int \mathrm{d}t \exp(\mathrm{i}t\mu), \qquad (3.2.8)$$

$$|\mu| = -\frac{1}{\pi} \fint \frac{\mathrm{d}s}{s} \partial_s \exp(\mathrm{i}s\mu), \qquad (3.2.9)$$

(where $\fint$ represents a Cauchy principal value integral with pole at 0). Where $\mathbf{t} = (t_x, t_y)$, equation (3.2.7) may be written in integral form

$$d_C = -\frac{1}{4\pi^3} \fint \frac{\mathrm{d}s}{s} \partial_s \int \mathrm{d}^2\mathbf{t} \langle \exp(\mathrm{i}\mathbf{t} \cdot \mathbf{J} + \mathrm{i}s \det \mathbf{M_{J}}_{\mathrm{sym}}) \rangle. \qquad (3.2.10)$$

The average is taken with respect to 12 gaussian random functions

$$\xi, \eta, \nabla\xi, \nabla\eta, \xi_{xx}, \xi_{yy}, \xi_{xy}, \eta_{xx}, \eta_{yy}, \eta_{xy};$$

of these, $(\xi, \xi_{xx}, \xi_{xy})$ and $(\eta, \eta_{xx}, \eta_{xy})$ are correlated. The first few integrals are straightforward gaussians, and in turn are taken with respect to $\nabla\xi, \nabla\eta, \mathbf{t}, \xi_{xy}, \eta_{xy}$, and then $(\xi_{xx}, \xi_{yy}, \eta_{xx}, \eta_{yy})$ as a vector. When this has been done, $\xi + i\eta$ is replaced with $\rho \exp(i\chi)$ and s is rescaled to $u = is$, yielding

$$d_C = -\frac{2}{\pi^3 K_2} \int_0^{2\pi} \mathrm{d}\chi \int_0^\infty \mathrm{d}\rho\, \rho \exp(-\rho^2/2) \int \frac{\mathrm{d}u}{u} \partial_u \frac{1}{(4 + iK_4 u)\sqrt{2 - i(K_4 - K_2^2)u}}. \tag{3.2.11}$$

The ρ, χ integrals are trivial, giving 4π, leaving only the Cauchy principal value integral in u. This may be evaluated as the average of two contour integrals, with the contour displaced in the complex plane both above and below the origin, each of which may now be safely integrated by parts (since neither contour now intersects the pole). The integrand now has a double pole at the origin, a simple pole in the upper half-plane at $u = 4i/K_4$, and a branch point in the lower half-plane at $-2i/(K_4 - K_2^2)$; the branch cut is taken along the imaginary axis to $-i\infty$. The upper contour can be deformed about the simple pole, giving $-\pi K_4^{3/2}/16\sqrt{3K_4 - 2K_2^2}$; the lower, deformed about the branch cut, can be integrated by elementary means to give $-\pi(K_4^{3/2}/\sqrt{3K_4 - 2K_2^2} - K_2^2)/16$. The final result is

$$d_C = \frac{K_4^{3/2}}{2\pi K_2 \sqrt{3K_4 - 2K_2^2}} - \frac{K_2}{4\pi}. \tag{3.2.12}$$

Since the dislocation density is $K_2/4\pi$, the total number of current zeros is

$$d_Z = \frac{K_4^{3/2}}{2\pi K_2 \sqrt{3K_4 - 2K_2^2}}. \tag{3.2.13}$$

Spatial averages ar replaced by ensemble averages in (2.2.8) to give the average Poincaré index d_I,

$$d_I = \langle \delta^2(\mathbf{J}) \det \mathbf{M_J} \rangle. \tag{3.2.14}$$

The determinant $\det \mathbf{M_J}$ again separates into contributions from the symmetric and antisymmetric parts, the latter being simply the dislocation density. Each of the summands

in (2.2.10) can be averaged in (3.2.14) separately:

$$\langle \delta^2(\mathbf{J}) \frac{1}{2} (\nabla \cdot \mathbf{J})^2 \rangle = \frac{K_4 - K_2^2}{2\pi K_2}$$

$$\langle \delta^2(\mathbf{J})(\xi \eta_{xy} - \eta \xi_{xy})^2 \rangle = \frac{K_4}{8\pi K_2}$$

$$\langle \delta^2(\mathbf{J}) \frac{1}{2} (\xi \eta_{xx} - \eta \xi_{xx})^2 \rangle = \langle \delta^2(\mathbf{J}) \frac{1}{2} (\xi \eta_{yy} - \eta \xi_{yy})^2 \rangle$$

$$= \frac{3K_4 - 2K_2^2}{16\pi K_2}, \qquad (3.2.15)$$

the result being that

$$d_I = 0. \qquad (3.2.16)$$

This implies that, just as topological charge is statistically neutral, so is the Poincaré index; this ought to be no surprise given that the field ψ is statistically stationary and invariant with respect to change of overall phase.

By equation (2.2.13), the density of phase saddles d_S is

$$d_S = \frac{1}{2}(d_Z - d_I) = \frac{d_Z}{2}$$

$$= \frac{K_4^{3/2}}{4\pi K_2 \sqrt{3K_4 - 2K_2^2}}, \qquad (3.2.17)$$

and by (2.2.14), the number of phase extrema is

$$d_E = d_C - d_S$$

$$= \frac{K_4^{3/2}}{4\pi K_2 \sqrt{3K_4 - 2K_2^2}} - \frac{K_2}{4\pi}. \qquad (3.2.18)$$

The densities of maxima and minima adding to give (3.2.18) are obviously equal. The size of the ratio of dislocations to extrema depends on the relative size of the fourth moment K_4 compared with the second moment K_2; if it is very small, most positive index current zeros are dislocations, but as K_4 dominates K_2, more critical points are extrema, as seen in the table.

The results for the five different spectra are tabulated in table 3.1, where densities of dislocations and saddles (extrema density is just the difference of these) are given for the various spectra, as well as a numerical approximation to the fraction f of dislocation density divided by density of positive index current zeros ($= d_S = d_D + d_E$).

The corresponding calculations for critical points in the intensity were made by [WH82], and are shown there to be rather more complicated than the relatively simple functions

	Dislocation density d_D	Saddle density d_S	$f = d_D/d_S$
General value	$\frac{K_2}{4\pi}$	$\frac{K_4^{3/2}}{4\pi K_2\sqrt{3K_4-2K_2^2}}$	-
Ring spectrum (units $\Lambda_{\mathrm{m}}^{-2}$)	π	π	1
Disk spectrum (units $\Lambda_{\mathrm{disk}}^{-2}$)	$\pi/2$	$\pi(2/3)^{3/2}$	0.919
Gaussian spectrum (units Λ_{σ}^{-2})	2π	$2\sqrt{2}\pi$	0.707
Shell spectrum (units $\lambda_{\mathrm{m}}^{-2}$)	$2\pi/3$	$2\sqrt{3}\pi/5$	0.962
Planck spectrum (units λ_{T}^{-2})	$80\pi^3/63$	$1764\sqrt{3}\pi^3/25\sqrt{2929}$	0.566

Table 3.1: A comparison of dislocation and saddle densities in the plane, for the five different spectra.

found here. Equation (3.2.11) shows that critical points are evenly distributed in phase χ, and negative exponentially distributed in intensity ρ^2, therefore being most likely to be found for low intensities. Therefore saddles and extrema are most likely to be found near dislocations, as would certainly be expected near dislocation creations/annihilations by conservation of Poincaré index (via the mechanisms (2.2.5), (2.2.6)), and agrees with (numerical) observations [SF95, Fre98a]. The probable proximity of saddles, extrema and dislocations implies that there is some form of index screening analogous to the dislocation charge screening discussed in the next section, but calculations for critical point correlation functions (involving a 24×24 correlation matrix) have not been possible.

3.2.3 Dislocation correlation functions

In this section, we consider the spatial structure of the distribution of dislocations, by calculating the (planar) two-point dislocation correlation functions, for both dislocation number and topological charge. The two points taken are $\mathbf{R}_A, \mathbf{R}_B$, and we use the field autocorrelation function $C(R)$ defined in equation (3.1.15). Isotropy and stationarity ensure that any two points may be taken, and it is only the distance between the two points that matters.

The *dislocation number correlation function* $g(R)$ is defined to be the (normalised) average of the product of two local densities, separated by $R = |\mathbf{R}_B - \mathbf{R}_A|$:

$$g(R) = \frac{\langle \delta(\xi_A)\delta(\eta_A)|\omega_A|\delta(\xi_B)\delta(\eta_B)|\omega_B|\rangle}{\langle \delta(\xi)\delta(\eta)|\omega|\rangle^2}. \tag{3.2.19}$$

As $R \to \infty$, it is expected that the two local densities become statistically independent, and $g(R) \to 1$ by the normalisation. The *dislocation charge correlation function* $g_Q(R)$ is defined to be the corresponding correlation for topological charge, where the local densities are weighted by $\text{sign}\,\omega$,

$$g_Q(R) = \frac{\langle \delta(\xi_A)\delta(\eta_A)\omega_A\delta(\xi_B)\delta(\eta_B)\omega_B\rangle}{\langle \delta(\xi)\delta(\eta)|\omega|\rangle^2}. \tag{3.2.20}$$

and, as $R \to \infty$, we expect $g_Q(R) \to 0$ by (3.2.4). We shall neglect any self-correlation (usually described by a δ-function at the origin).

If there is no correlation, the local dislocation densities are completely independent for all R; this is the situation for a random distribution of signed points (Poisson distribution) in the plane, which can have any specified density. Such a set of Poisson distributed dots clearly have

$$g_{\text{Poisson}}(R) = 1, \qquad g_{Q\,\text{Poisson}}(R) = 0. \tag{3.2.21}$$

Comparison shall also be made to another situation of arrangements of signed points in the plane (rather than continuous field autocorrelations), namely the case of ionic liquids, where, for simplicity, we consider two species (ions) identical except for opposite charges ± 1, for example as considered by [HM86, HM75]. In liquid theory, the partial correlation functions $g_{++}(R), g_{+-}(R)$ are frequently studied, where

$$g_{++}(R) \equiv g(R) + g_Q(R), \qquad g_{+-}(R) \equiv g(R) - g_Q(R). \tag{3.2.22}$$

Both of these functions approach 1 as $R \to \infty$, and represent the correlation of dislocations with others of the same sign (different sign) alone.

If there are two dislocations close together (R is small), then expanding ψ, generically the dislocations are of opposite signs, and in the limit $R \to 0$,

$$g(0) = -g_Q(0), \tag{3.2.23}$$

implying $g_{++}(0) = 0$, so there is a statistical repulsion of like-signed dislocations, a statistical effect of the higher codimension of high-strength singularities. Note that all references

to attraction, repulsion, etc are statistical; no dynamics are implicit in the correlation functions $g(R), g_Q(R)$, only the average over the ensembles of wavefields ψ.

The calculations involve the equations (3.1.19), and for convenience the notations

$$
\begin{aligned}
C &\equiv C(R) \\
E &\equiv C'(R) \\
H &\equiv -C'(R)/R \\
F &\equiv -C''(R) \\
F_0 &\equiv -C''(0) = K_2/2 \qquad (3.2.24)
\end{aligned}
$$

are used, recalling that $C(0) = 1$ by normalisation. The displacement R, is taken to be in the x-direction for the calculation, as described in section 3.1. This implies that in the averages (3.2.19), (3.2.20), there are two different nondiagonal correlation matrices, with the two gaussian random vectors in ξ (and equivalently in η)

$$
\mathbf{u}_1 = (\xi_A, \xi_B, \xi_{Ax}, \xi_{Bx}), \qquad \mathbf{u}_2 = (\xi_{Ay}, \xi_{By}), \qquad (3.2.25)
$$

with correlation matrices

$$
\mathbf{M}_1 = \begin{pmatrix} 1 & C & 0 & E \\ C & 1 & -E & 0 \\ 0 & -E & F_0 & F \\ E & 0 & F & F_0 \end{pmatrix}, \qquad \mathbf{M}_2 = \begin{pmatrix} F_0 & H \\ H & F_0 \end{pmatrix}. \qquad (3.2.26)
$$

These matrices have determinants

$$
\begin{aligned}
\det \mathbf{M}_1 \equiv D_1 &= [E^2 - (1+C)(F_0 - F)][E^2 - (1-C)(F_0 + F)], \\
\det \mathbf{M}_2 \equiv D_2 &= F_0^2 - H^2, \qquad (3.2.27)
\end{aligned}
$$

with relevant probability densities

$$
\begin{aligned}
P(\xi_A = 0, \xi_B = 0, \xi_{Ax}, \xi_{Bx}) &= \frac{\exp(-\mathbf{u}_1' \cdot \mathbf{N}_1 \cdot \mathbf{u}_1'/2)}{4\pi^2 \sqrt{D_1}}, \\
P(\xi_{Ay}, \xi_{By}) &= \frac{\exp(-\mathbf{u}_2 \cdot \mathbf{N}_2 \cdot \mathbf{u}_2/2)}{2\pi \sqrt{D_2}}, \qquad (3.2.28)
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbf{u}_1' &\equiv (\xi_{Ax}, \xi_{Bx}) \\
\mathbf{N}_1 &= \frac{1}{D_1} \begin{pmatrix} -[E^2 - F_0(1-C^2)] & CE^2 - F(1-C^2) \\ CE^2 - F(1-C^2) & -[E^2 - F_0(1-C^2)] \end{pmatrix} \\
\mathbf{N}_2 &= \frac{1}{D_2} \begin{pmatrix} F_0 & -H \\ -H & F_0 \end{pmatrix}
\end{aligned}
\tag{3.2.29}
$$

and similarly for η, for which the correlation vectors are written $\mathbf{v}$ instead of $\mathbf{u}$ (and noting the slight corrections from [BD00] equations (4.39), (4.40)).

The number correlation function $g(R)$ is rather complicated to evaluate, mainly due to the presence of the modulus signs in ω; these are transformed using a Fourier identity similar to (3.2.9),

$$
|\omega| = \frac{1}{\pi} \int \frac{\mathrm{d}t}{t^2} (1 - (\exp(\mathrm{i}\omega t) + \exp(-\mathrm{i}\omega t))/2). \tag{3.2.30}
$$

Applying this to (3.2.19) and multiplying out, one finds

$$
g(R) = \frac{1}{d_D^2 \pi^2} \int \frac{\mathrm{d}t_A}{t_A^2} \int \frac{\mathrm{d}t_B}{t_B^2} [T(0,0) - T(t_A, 0) - T(0, t_B) + (T(t_A, t_B) + T(t_A, -t_B))/2], \tag{3.2.31}
$$

where

$$
T(t_A, t_B) = \langle \delta(\xi_A) \delta(\eta_A) \delta(\xi_B) \delta(\eta_B) \exp(\mathrm{i}(\omega_A t_A - \omega_B t_B)) \rangle. \tag{3.2.32}
$$

Since the ω are quadratic in the integration variables, each such T is a gaussian vector integral, which is easy to evaluate, using (3.2.28), (3.2.29),

$$
T(t_A, t_B) = \frac{1}{(2\pi)^2 D(t_A, t_B)}, \tag{3.2.33}
$$

where $D(t_A, t_B)$ is the determinant of the matrix in the gaussian integral, given by

$$
D(t_A, t_B) = (1-C^2) + (t_A^2 + t_B^2) F_0(E^2 - F_0(1-C^2)) - t_A t_B H(CE^2 - F(1-C^2)) + t_A^2 t_B^2 D_1 D_2. \tag{3.2.34}
$$

t_A, t_B are now rescaled, and defining the R-dependent functions

$$
\begin{aligned}
Y &\equiv \frac{H^2(CE^2 - F(1-C^2))^2}{F_0^2(E^2 - F_0(1-C^2))^2} \\
Z &\equiv \frac{D_1 D_2(1-C^2)}{F_0^2(E^2 - F_0(1-C^2))^2} \\
X &\equiv \frac{[E^2 H - F_0(CE^2 - (1-C^2)(F-H))][E^2 H + F_0(CE^2 - (1-C^2)(F+H))]}{F_0^2(E^2 - F_0(1-C^2))^2} \\
&= \sqrt{(1 - Y + Z)^2 - 4Z}
\end{aligned}
\tag{3.2.35}
$$

the expression becomes

$$g(R) = \frac{F_0(E^2 - F_0(1-C^2))}{4\pi^4 d_D^2 (1-C^2)^2} \int \frac{\mathrm{d}t_A}{t_A^2} \int \frac{\mathrm{d}t_B}{t_B^2} I(t_A, t_B, Y, Z), \qquad (3.2.36)$$

where

$$I(t_A, t_B, Y, Z) = 1 - \frac{1}{1+t_A^2} - \frac{1}{1+t_B^2} + \frac{1 + t_A^2 + t_B^2 + Z t_A^2 t_B^2}{(1 + t_A^2 + t_B^2 + Z t_A^2 t_B^2)^2 - 4Y t_A^2 t_B^2}. \qquad (3.2.37)$$

t_A is now evaluated by residues, and this leaves the one remaining integral

$$g(R) = -\frac{2(E^2 - F_0(1-C^2))}{\pi F_0 (1-C^2)^2} \int_0^\infty \frac{3 - Z + 2Y + (3 + Z - 2Y)t^2 + 2Z t^4}{(1+t^2)\sqrt{1 + (1+Z-Y)t^2 + Z t^4}}. \qquad (3.2.38)$$

It is possible to evaluate this integral, and the result involves elliptic functions in a rather nonilluminating way;[1]

$$g(R) = \frac{2(E^2 - F_0(1-C^2))}{\pi F_0 (1-C^2)^2} \Big(2\sqrt{2 - Y + 2Z} - \frac{\mathrm{i}}{\sqrt{2UZ}} [(4-U)Z F_p - 4Z E_p$$
$$+ 2YU\Pi_p + 2\sqrt{Z}(U E_m + 2Y\Pi_m - (1 + X + Y)F_m)] \Big), \qquad (3.2.39)$$

where

$$\begin{aligned}
F_p &= F(\mathrm{i}\operatorname{arcsinh}(\sqrt{V/2})|U/V), \\
F_m &= F(-\mathrm{i}\operatorname{arcsinh}(\sqrt{2/V})|V/U), \\
E_p &= E(\mathrm{i}\operatorname{arcsinh}(\sqrt{V/2})|U/V), \\
E_m &= E(-\mathrm{i}\operatorname{arcsinh}(\sqrt{2/V})|V/U), \\
\Pi_p &= \Pi(2/V; \mathrm{i}\operatorname{arcsinh}(\sqrt{V/2})|U/V), \\
\Pi_m &= \Pi(V/2; -\mathrm{i}\operatorname{arcsinh}(\sqrt{2/V})|V/U)
\end{aligned} \qquad (3.2.40)$$

where F, E, Π are the incomplete elliptic functions of the first, second and third kinds respectively (with the conventions for elliptic functions being those of *Mathematica* [Wol99]) and the symbols used are not to be confused with other definitions of E, F, Π. Finally,

$$U \equiv 1 + X - Y + Z, \qquad V \equiv 1 - X - Y + Z. \qquad (3.2.41)$$

The charge correlation function $g_Q(R)$ is considerably easier to find than the number correlation $g(R)$; having integrated out the trivial δ-functions, the eight first derivatives

---

[1]The integral (3.2.38) was calculated symbolically using Mathematica, with (3.2.39) being the (simplified) output. It is possible this function may be simplified further. See [SBS01] for an alternative derivation.

remain, giving

$$g_Q(R) = \frac{1}{d_D^2 (2\pi)^6 D_1 D_2} \int d^2 \nabla \xi_A d^2 \nabla \xi_B d^2 \nabla \eta_A d^2 \nabla \eta_B \, \omega_A \omega_B$$

$$\times \exp(-(\mathbf{u}_1' \cdot \mathbf{N}_1 \cdot \mathbf{u}_1' + \mathbf{u}_2 \cdot \mathbf{N}_2 \cdot \mathbf{u}_2 + \mathbf{v}_1' \cdot \mathbf{N}_1 \cdot \mathbf{v}_1' + \mathbf{v}_2 \cdot \mathbf{N}_2 \cdot \mathbf{u}_2)/2). \quad (3.2.42)$$

The gaussians are Fourier transformed, and, integrating by parts, the quadratic terms in ω_A, ω_B can be replaced with derivatives. The resulting the integration is easy, producing

$$g_Q(R) = \frac{2E(CE^2 - F(1 - C^2))}{RF_0^2 (1 - C^2)^2} = \frac{1}{C_0''^2 R} \partial_R \left[\frac{C'(R)^2}{1 - C(R)^2} \right] \quad (3.2.43)$$

The first equation of (3.2.43) is a special case of equation (6.27) in [Hal81].

For the five spectra, the number and charge correlation functions $g(R), g_Q(R)$ are shown in figure (3.6), and the partial correlation functions $g_{++}(R), g_{+-}(R)$ in figure (3.7), where the repulsion of like charges is manifest. When $R = 0$,

$$g(0), g_Q(0) = \pm \frac{1}{2} \left(\frac{C(0)''''}{C(0)''^2} - 1 \right) = \pm \frac{1}{2} \left(\frac{3K_4}{2K_2^2} - 1 \right), \quad (3.2.44)$$

the same combination of K_2, K_4 which appeared in the phase critical point calculations. The value of $g(0), g_Q(0)$ is inversely related to the ratio $f = d_D/d_S$ of dislocations to saddles, listed for the five spectra in table 3.1. Dislocations (of opposite sign) are obviously more likely to be closer together in waves whose spectra have many short waves.

The second equation of (3.2.43) implies the relation

$$2\pi d_D \int_0^\infty dR \, R g_Q(R) = -1 \quad (3.2.45)$$

which is a *screening relation*, saying that, on average, the integral of the topological charge throughout the rest of the plane must compensate the strength of a dislocation centred on the origin (whose self interaction we neglect). In ionic structure theory, (3.2.45) is known as the first Stillinger-Lovett sum rule [SL68b, SL68a, HM86], in the particularly simple case of two equal but oppositely charged species. The second Stillinger-Lovett sum rule related the second moment of R with respect to g_Q with the screening length, that is, the effective distance over which the screening in (3.2.45) takes place. It is perhaps striking that these tools of statistical mechanics, conceived to handle situations where there is a pairwise coulombic interaction between the points, should apply here, to the zeros of complex wavefields signed by the topological charge. The generalisation of these rules to zeros of the chaotic analytic function [Han98c] was observed by Hannay [Han96],
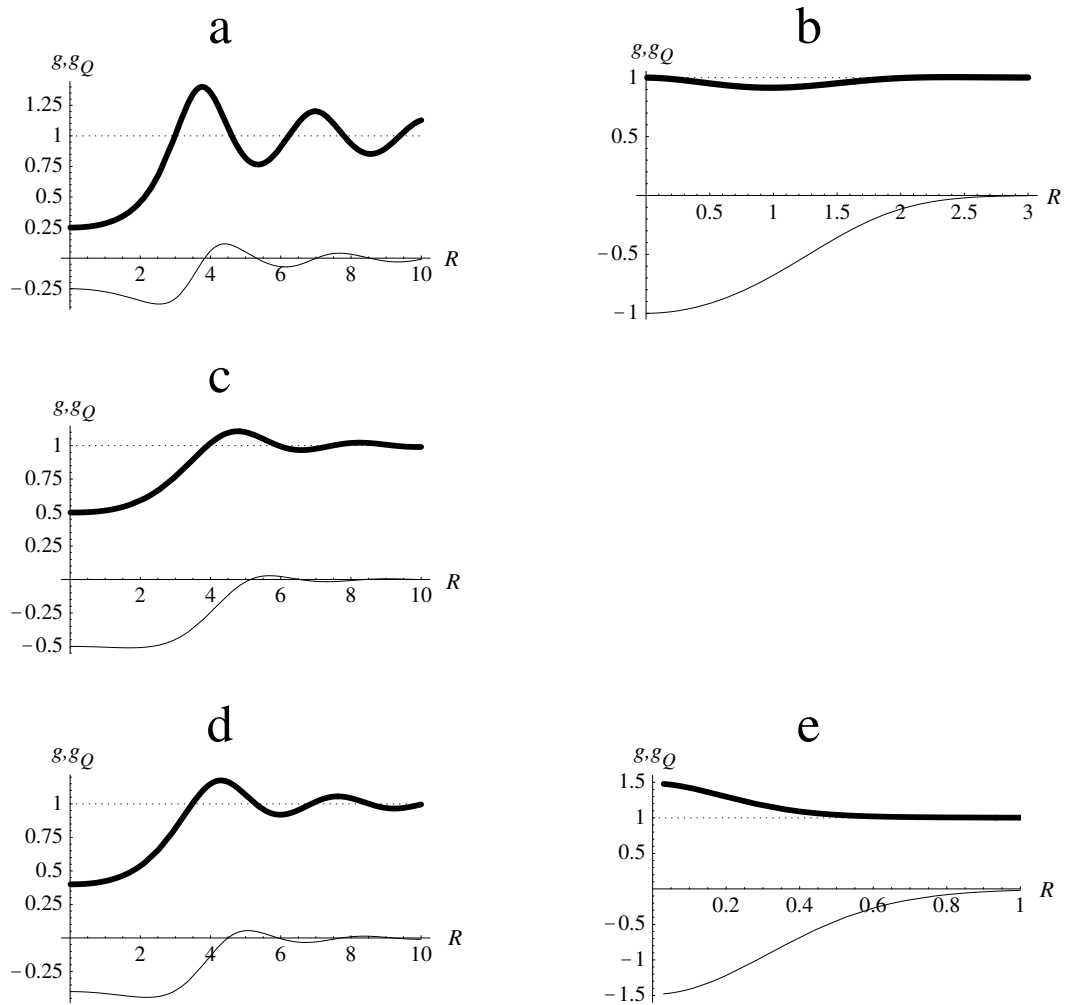
Figure 3.6: Number and charge correlation functions $g(R), g_Q(R)$ for the five spectra, in units of the appropriate inverse wavenumber. The thicker line represents $g(R)$, the thinner $g_Q(R)$. (a) Ring spectrum, (b) Gaussian spectrum, (c) Disk spectrum, (d) Shell spectrum, (e) Planck spectrum.
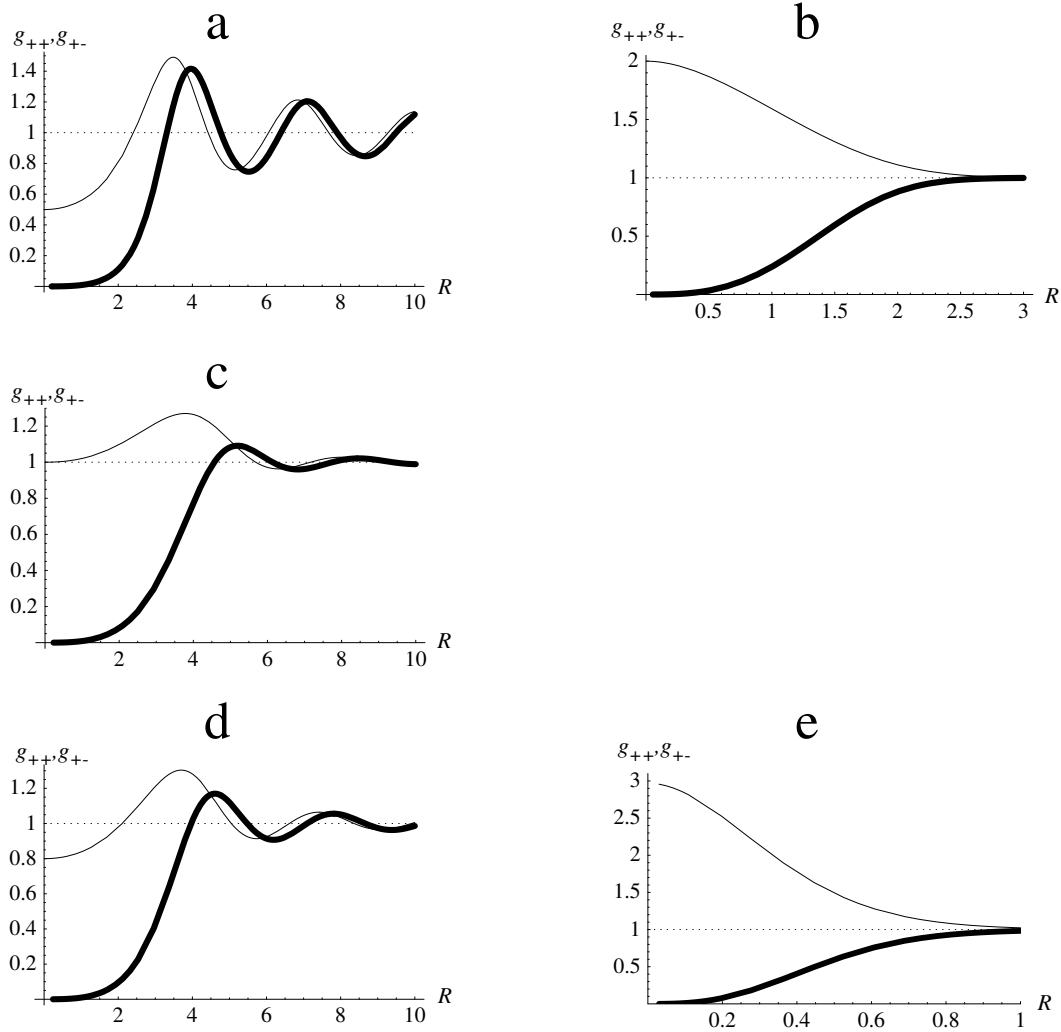
Figure 3.7: Partial correlation functions $g_{++}(R), g_{+-}(R)$ for the five spectra, in units of the appropriate inverse wavenumber. The thicker line represents $g_{++}(R)$, the thinner $g_{+-}(R)$.(a) Ring spectrum, (b) Gaussian spectrum, (c) Disk spectrum, (d) Shell spectrum, (e) Planck spectrum.

from coulombic calculation of Jancovici [Jan87] although in these cases there was only one species.

The nature of the screening may be understood by considering the charge $Q(N)$ associated with the dislocations in an area $\mathcal{A} = N/d_D$, where the mean number of dislocations in $\mathcal{A}$ is N ($\gg 1$). The average charge $\langle Q(N) \rangle$ is obviously zero, but what about its mean square fluctuation $\langle Q^2(N) \rangle$? If the charges merely have average neutrality, then there would be no long-range correlations, and we would expect $\langle Q^2(N) \rangle \sim N$. However, the nontrivial correlation functions give a more interesting result; the boundary of $\mathcal{A}$ must be smoothed (here by a gaussian) in order that the subtle behaviour is not dominated by unimportant edge effects (this was not done in [FW98]). With the circular $\mathcal{A}$ gaussian weighted, we have

$$
\begin{aligned}
\langle Q^2(N) \rangle &= \frac{1}{2}N \left(1 + 2\pi d_D \int_0^\infty \mathrm{d}R\, R g_Q(R) \exp\left(-\frac{\pi R^2}{2\mathcal{A}} \right) \right) \\
&= \frac{1}{4} \int_0^\infty \mathrm{d}R\, R \frac{C'(R)^2}{1 - C(R)^2} \exp\left(-\frac{\pi R^2}{2\mathcal{A}} \right)
\end{aligned}
\tag{3.2.46}
$$

where (3.2.43), (3.2.45) have been used. The first equality shows that, without screening (as in the Poisson case where $g_Q(R) = 0$), the leading term for large N would be $N/2$, and the fluctuations would be those of a random distribution with overall neutrality. However, for dislocations there is screening, and expanding in $1/N$,

$$
\langle Q^2(N) \rangle = \frac{1}{4} \int_0^\infty \mathrm{d}R\, R \left[\frac{C'(R)^2}{1 - C(R)^2} \right] + \mathcal{O}(N^{-1}),
\tag{3.2.47}
$$

provided the integral converges, leaving fluctuations that are *independent of N* for large N. For the sharp spectra (the ring spectrum (3.1.27), shell spectrum (3.1.36)), the integral does not converge, and asymptotics on (3.2.46) show that $\langle Q^2(N) \rangle \sim \log N$ for the shell spectrum, and $\langle Q^2(N) \rangle \sim \sqrt{N}$ for the ring spectrum.

The relationship between dislocation correlation functions and salt correlation functions must be treated with caution. Although the sharp spectra have oscillations in their correlation functions (figures (3.6,3.7a,c,d)), the physical origin of the oscillation is different. For molten salts, the oscillation originates simply from packing effects due to the finite radii of the charged species; for dislocations, packing is no problem, but any sharp cutoff in the spectrum Π leads to oscillations in the correlation function $C(R)$ (the Fourier transform of the spectrum), and, excepting the disk spectrum, it is these spectra which have infinite screening length. The ring spectrum in particular has long range structure;

the J_0 correlation function decays very slowly, and the geometric feature on the large scale of these fields is very striking, and discussed in [OGH87]. It is possible that the 'scarlet' structures observed here may be quantified using dislocation correlations. Such patterns are not seen for the other spectra.

The fact that the oscillations in the correlation function are not due to packing is made more obvious when one looks at the partial correlation functions in figure (3.7a,c,d): the oscillations of g_{++}, g_{+-} are almost in phase, whereas one expects them to be out of phase from packing considerations [HM86]. The disk spectrum, with oscillations yet a finite screening length, gives the closest comparison with ionic salts, but even its partial correlation functions (plotted in figure (3.7c)) are in phase. There does not seem to be any analogue in ionic theory to the correlation functions of the Gaussian and Planck spectra (figure (3.6b,e)), where there are no oscillations at all.

The nearest neighbour probability density functions $P(R), P_{++}(R), P_{+-}(R)$ may be crudely estimated from the two-point correlation functions $g(R), g_{++}(R), g_{+-}(R)$ by a Poisson approximation (neglecting the correlations for between more numbers of points). The probability that the nearest neighbour is a distance R from the singularity is taken to be the probability of there being no dislocations in the annular rings of radii $R_1, R_2, \ldots,$ R_{M-1}, with small widths δR_i, and there being one dislocation in the ring with radius $R = R_M$, width δR, the approximation being that the number in each such annular ring is independent of the others (the true probability calculation would involve all multipoint correlation functions). With this approximation,

$$P(R)\mathrm{d}R = 2\pi R\,\delta R\,d_D g(R) \prod_{i=1}^{M-1} (1 - 2\pi R_i \delta R_i\, d_D g(R_i)), \qquad (3.2.48)$$

and in the limit $M \to \infty$, $\delta R_i \to 0$,

$$P(R) = 2\pi R d_R g(R) \exp\left(-2\pi d_D \int_0^R \mathrm{d}R'\, R' g(R')\right), \qquad (3.2.49)$$

with appropriate substitutions for the partial correlation functions. For randomly distributed (Poisson) points, $g(R) = 1$ and (3.2.49) is simply a two-dimensional Poisson distribution (ie a Rayleigh distribution), whose length scale is fixed by $1/\sqrt{d_D}$. The distributions for the different spectra is given in figure (3.8), along with the earlier $1/\sqrt{d_D}$ estimate for mean spacing (providing the scale for the corresponding Poisson distribution). The repulsion of like charges is again clearly represented. A similar calculation
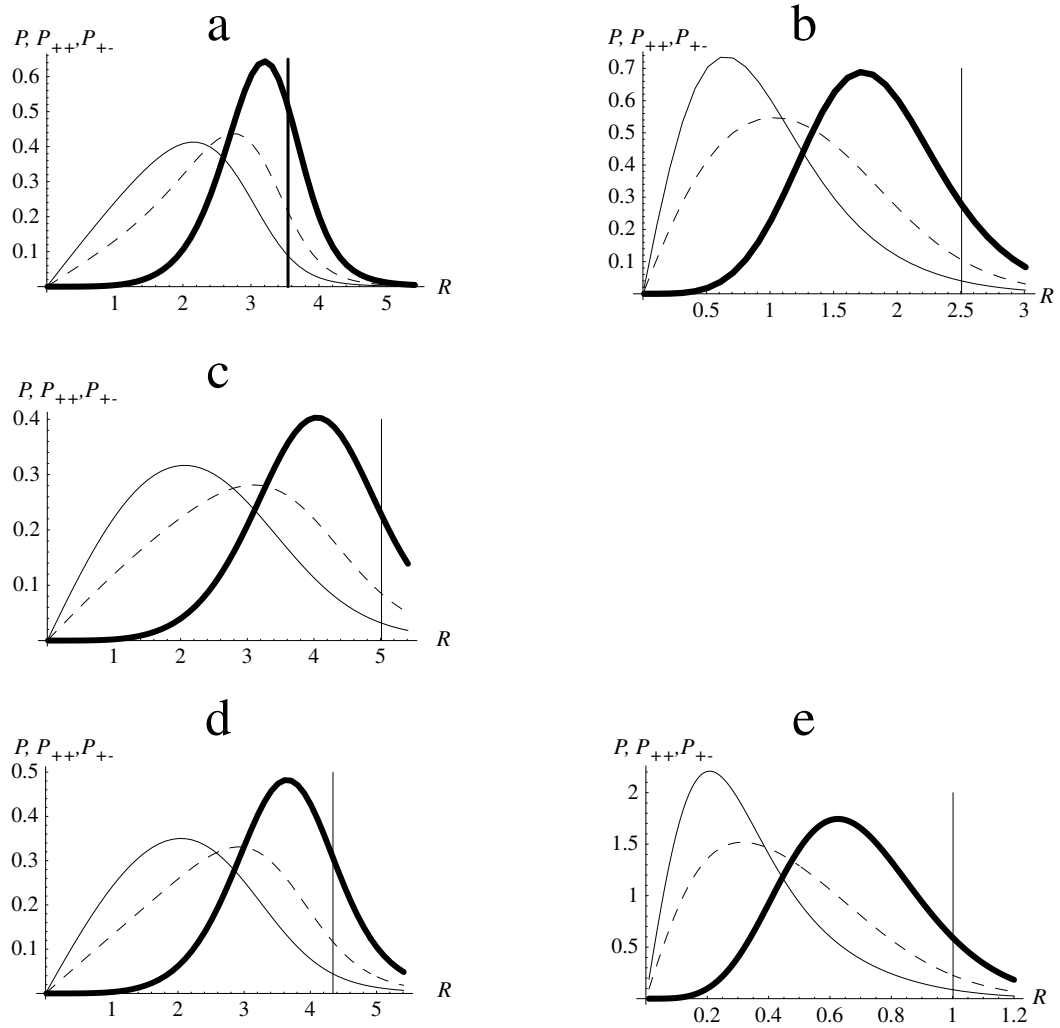
Figure 3.8: Approximate nearest neighbour probability densities $P(R)$ (dashed line), $P_{++}(R)$ (thick line), $P_{+-}(R)$ (thin line) for the five spectra, (with R in units of inverse wavenumber): (a) ring; (b) gaussian; (c) disk; (d) shell; (e) Planck.

(and refinement based on the Bernoulli approximation), was made for the ring spectrum by [SBS01], who also found good agreement between theory and simulation.

3.2.4   Phase anisotropy ellipse eccentricity probability density

Hitherto only densities of singularities and their correlations have been calculated; here we begin calculating the averages of geometric quantities associated with dislocations, in particular the distribution of eccentricity of the anisotropy ellipse, described in section 2.3, with eccentricity ε given in terms of G, ω in equation (2.3.6). It is easier to manipulate ε^2,

$$\varepsilon^2 = \frac{1}{2\omega^2}\sqrt{G^2 - 4\omega^2}\left(G - \sqrt{G^2 - 4\omega^2}\right), \tag{3.2.50}$$

where the joint probability of G and ω was given in equation (3.1.26). The probability distribution of ε can be found in terms of ε^2, using

$$P(\varepsilon) = \frac{\mathrm{d}\varepsilon^2}{\mathrm{d}\varepsilon}P(\varepsilon^2) = 2\varepsilon P(\varepsilon^2). \tag{3.2.51}$$

The probability distribution of ε^2 for dislocations is therefore (using equations (3.1.24), (3.1.26), (3.2.5))

$$
\begin{aligned}
P(\varepsilon^2) &= \left\langle \delta\left(\varepsilon^2 - \frac{1}{2\omega^2}\sqrt{G^2 - 4\omega^2}\left(G - \sqrt{G^2 - 4\omega^2}\right)\right)\right\rangle_{\mathrm{d}} \\
&= \frac{4}{K_2^3}\int_0^\infty \mathrm{d}G\,\exp(-G/K_2)\int_0^{G/2}\mathrm{d}\omega \\
&\qquad \times \omega\delta\left(\varepsilon^2 - \frac{1}{2\omega^2}\sqrt{G^2 - 4\omega^2}\left(G - \sqrt{G^2 - 4\omega^2}\right)\right) \\
&\qquad\qquad (\text{rescaling } \omega \to u = 2\omega/G) \\
&= \frac{4}{K_2^3}\underbrace{\int_0^\infty \mathrm{d}G\,(G^2/4)\exp(-G/K_2)}_{K_2^3/2}\int_0^1 \mathrm{d}u\, u\delta(\varepsilon^2 - \frac{2}{u^2}\sqrt{1 - u^2}(1 - \sqrt{1 - u^2})) \\
&\qquad\qquad (\text{rescaling } u \to v = \sqrt{1 - u^2}) \\
&= 2\int_0^1 \mathrm{d}v\, v\delta(\varepsilon^2 - 2v(1 - v)/(1 - v^2)) \\
&\qquad\qquad (\text{rescaling } v \to w = 2v/(1 + v)) \\
&= 4\int_0^1 \mathrm{d}w\,\frac{w}{(2 - w)^3}\delta(\varepsilon^2 - w) \\
&= \frac{4\varepsilon^2}{(2 - \varepsilon^2)^3}. \tag{3.2.52}
\end{aligned}
$$

Thus, by (3.2.51), the probability density for the eccentricity of the core anisotropy ellipse,

in two dimensions, is

$$P_2(\varepsilon) = \frac{8\varepsilon^3}{(2 - \varepsilon^2)^3}. \tag{3.2.53}$$

The moments of this distribution are

$$\langle \varepsilon \rangle_{d,2} = \frac{8}{4 + n} \, _2F_1(1, 3; (3 + n)/2; -1), \tag{3.2.54}$$

where $_2F_1$ denotes the (Gauss) hypergeometric function [AS65]. The first two moments are

$$\langle \varepsilon \rangle_{d,2} = \operatorname{arctanh} \frac{1}{\sqrt{2}} - 1 \approx 0.8697$$

$$\langle \varepsilon^2 \rangle_{d,2} = 2(2 \log 2 - 1) \approx 0.7726 \tag{3.2.55}$$

The probability distribution and moments are all completely independent of the spectrum, unlike the other averages (density, speed, etc). This is because the eccentricity only depends on quadratic objects involving first derivatives (G and ω), giving rise to a factor of K_2 which is cancelled in the dislocation normalisation. The average eccentricity $\langle \varepsilon \rangle_{d,2}$ is quite large and universal, implying that a generic dislocation is in fact rather anisotropic. An ellipse with this eccentricity (with phase lines) is shown in figure (3.9). Note that this eccentricity is not the same, in general, as the eccentricity on average of the ellipse associated with $\nabla \psi$ at an arbitrary point in the field, since the first line of (3.2.52) involves the dislocation average (3.2.5), with an extra factor of $|\omega|$ in the integral.

3.2.5 Planar speed probability density

The final planar statistic we consider is that of speed V, given in equation (2.4.4), which involves the averaging the gaussian random functions $\xi, \eta, \nabla \xi, \nabla \eta, \xi_t, \eta_t$. Therefore, the calculation involves the correlation matrices

$$\mathbf{M}_\pm = \begin{pmatrix} 1 & \pm cK_1 \\ \pm cK_1 & c^2 K_2 \end{pmatrix} \tag{3.2.56}$$

for the vectors $\mathbf{u}_+ = (\xi, \eta_t), \mathbf{u}_- = (\eta, \xi_t)$. Note that for plane sections of spatial waves, k_n should be used here rather than the projected K_n, as discussed after the equations (3.1.18). The matrices $\mathbf{M}_\pm$ imply that the relevant joint probability density functions (when $\xi, \eta = 0$) are

$$P(\xi = 0, \eta_t) = \frac{1}{2\pi v_c \sqrt{K_2}} \exp(-\eta_t^2/2v_c^2 k_2)$$

$$P(\eta = 0, \xi_t) = \frac{1}{2\pi v_c \sqrt{K_2}} \exp(-\xi_t^2/2v_c^2 k_2), \tag{3.2.57}$$

where v_c is the *characteristic velocity*, defined

$$v_c \equiv \begin{cases} c\sqrt{1 - K_1^2/K_2} & \text{(two-dimensional wave)} \\ c\sqrt{1 - k_1^2/k_2} & \text{(plane section of three-dimensional wave)} \end{cases} \quad . \quad (3.2.58)$$

v_c is limited by the speed of the carrier wave, reaching that limit as $K_1^2/K_2 \to 0$. V is rescaled to be in units of v_c,

$$V_{sc} \equiv V/v_c, \quad (3.2.59)$$

and the probability density function for speed is found using

$$P(V) = \frac{1}{v_c} P(V_{sc}) = \frac{2V}{v_c^2} P(V_{sc}^2). \quad (3.2.60)$$

Using equations (2.4.5), (3.1.24), we have

$$\begin{aligned} P(V_{sc}^2) &= \left\langle \delta \left(V_{sc}^2 - \left[\frac{(\xi_t^2 (\nabla \eta)^2 + \eta_t^2 (\nabla \xi)^2 - 2\xi_t \eta_t \nabla \xi \cdot \nabla \eta)}{v_c^2 \omega^2} \right] \right) \right\rangle_d \\ &= \frac{1}{2\pi} \int dq \, \exp(-iq V_{sc}^2) \left\langle \exp \left(iq \left[\frac{(\xi_t^2 (\nabla \eta)^2 + \eta_t^2 (\nabla \xi)^2 - 2\xi_t \eta_t \nabla \xi \cdot \nabla \eta)}{v_c^2 \omega^2} \right] \right) \right\rangle_d \\ &\qquad \text{(using equation (3.2.8))} \\ &= \frac{4}{\pi} \int dq \int_0^\infty dG \int_0^{G/2} d\omega \, \frac{\omega^2 \exp(-G + iq V_{sc}^2)}{\sqrt{\omega^2 + 2iGq - 4q^2}}, \quad (3.2.61) \end{aligned}$$

where in the final line, equations (3.1.26), (3.2.57), have been used, and G, ω rescaled. The final integral is evaluated (see [BD00] appendix B) to give

$$P(V_{sc}^2) = \frac{2}{(2 + V_{sc}^2)^2}, \quad (3.2.62)$$

which, by (3.2.60), implies

$$P(V) = \frac{4v_c^2 V}{(2v_c^2 + V^2)^2}. \quad (3.2.63)$$

Although v_c cannot be greater than the speed c of the component waves, the distribution (3.2.63) is not limited, and admits arbitrarily high speeds.

All moments of this distribution diverge except for the first, which is

$$\langle V \rangle_d = \frac{\pi v_c}{\sqrt{2}}. \quad (3.2.64)$$

The only spectrum we are considering with moving dislocations is the Planck spectrum, for which, by (3.1.42),

$$v_{c \, \text{Planck}} = c\sqrt{1 - \frac{68040 \zeta(5)^2}{\pi^{10}}} \approx 0.4678c, \quad (3.2.65)$$

so the average speed of dislocations in plane sections of (scalar) blackbody radiation is $1.039c$; faster, on average, than the speed of light.

3.3 Statistical geometry of dislocation lines in three dimensions

We now come to the calculations of statistical properties of dislocation lines in space, calculating quantities that were found in chapter 2 sections 2.5, 2.7, 2.8. As before, all quantities depend only on the spatial moments k_n of the spectrum (3.1.8).

3.3.1 Dislocation line density

The average length of dislocation line in space, $d_{D,3}$ plays an identical role for the three dimensional statistics as the two dimensional density d_D (3.2.1) did in two. By ergodicity, it may be found using the formula (2.5.1) for dislocation length in a volume, so

$$d_{D,3} = \langle \delta(\xi)\delta(\eta)|\nabla\xi \wedge \nabla\eta|\rangle, \tag{3.3.1}$$

where now the ensemble is that of the three dimensional waves (3.1.1). Using a similar trick of writing the (rescaled) coordinates of $\nabla\eta = \mathbf{Y}\sqrt{k_2/3}$ in terms of $\nabla\xi = \mathbf{X}\sqrt{k_2/3}$, (transforming the $\mathbf{X}$ direction to the z-axis gives jacobian 4π and $\mathbf{Y}$ has polar angles θ, ϕ with respect to this direction),

$$
\begin{aligned}
d_{D,3} &= \frac{1}{(2\pi)^4}\left(\frac{3}{k_2}\right)^3 \int d^3\nabla\xi d^3\nabla\eta \, |\nabla\xi \wedge \nabla\eta| \exp(-3((\nabla\xi)^2 + (\nabla\eta)^2)/2k_2) \\
&= \frac{4\pi k_2}{3(2\pi)^4} \int_0^\infty dX \int_0^\infty dY \int_0^{2\pi} d\phi \int_0^\pi d\theta \, X^3 Y^3 \sin^2\theta \exp(-(X^2 + Y^2)/2) \\
&= \frac{k_2}{3\pi}
\end{aligned}
\tag{3.3.2}
$$

Since geometric properties are invariant under rotation, we shall frequently choose a simple local frame when spherical averaging is necessary.

Comparison with (3.2.2) shows that, for a plane section of a three dimensional wave,

$$d_{D,2} = \frac{k_2}{6\pi} = \frac{d_{D,3}}{2}, \tag{3.3.3}$$

the ratio of $1/2$ coming about since

$$\frac{d_{D,2}}{d_{D,3}} = \frac{\langle|\omega_z|\rangle}{\langle|\boldsymbol{\omega}|\rangle}, \tag{3.3.4}$$

and the average modulus of the z-component of any isotropically distributed unit vector $\mathbf{e}$ is

$$\langle|e_z|\rangle = \frac{1}{4\pi}\int_0^{2\pi} d\phi \int_0^\pi d\theta \, \sin\theta|\cos\theta| = \frac{1}{2}. \tag{3.3.5}$$

This result holds generally for the distribution of points in plane sections of any random set of lines which are isotropically distributed, and shall be used in the corresponding calculations of three dimensional polarization singularity densities in section 4.6.

Dislocation averages in three dimensions are defined, in analogy with (3.2.5). For a quantity f, its dislocation average is

$$\langle f \rangle_{d,3} \equiv \frac{1}{d_{D,3}} \langle \delta(\xi)\delta(\eta)|\omega|f \rangle, \tag{3.3.6}$$

which again gives the correct statistical weighting.

3.3.2 Anisotropy ellipse probability density in three dimensions

As stated in 2.5, the anisotropy ellipse is defined in three dimensions in the $(\nabla\xi, \nabla\eta)$ plane normal to the dislocation direction. With the three dimensional definition of G, ω, distributed as in (3.1.25), the squared eccentricity ε^2 is the same as in equation (3.2.50). The calculation is very similar to its two dimensional analogue (3.2.52), and the eccentricity probability density is

$$P_3(\varepsilon) = \frac{24\varepsilon^3\sqrt{1-\varepsilon^2}}{(2-\varepsilon^2)^4}, \tag{3.3.7}$$

where, of course, $0 \leq \varepsilon \leq 1$. This distribution has moments

$$\langle \varepsilon^n \rangle = \frac{6\sqrt{\pi}(1+n/2)!}{((5+n)/2)!} \, {}_2F_1(3/2, 4; (7+n)/2; -1) \tag{3.3.8}$$

and the first two moments are given explicitly by

$$\langle \varepsilon \rangle_{d,3} = \frac{3\pi}{8\sqrt{2}} = 0.8330,$$

$$\langle \varepsilon^2 \rangle_{d,3} = \frac{(3\pi - 8)}{2} = 0.7124. \tag{3.3.9}$$

As with the two-dimensional case, these are universal for all spectra, but are less than the corresponding moments (3.2.55). This is expected, since if an elliptic cylinder is obliquely sliced (the dislocation crosses the plane, but is not normal to it), the eccentricity of the slice is greater than that of the transverse ellipse. Ellipses with the eccentricities $\langle \varepsilon \rangle_{d,2}, \langle \varepsilon \rangle_{d,3}$ are shown in figure (3.9).

3.3.3 Dislocation speed in three dimensions

The relationship between the distribution of speeds in two and three dimensions is similar to the distribution of ellipse eccentricities; the expressions for speed (2.4.4) and its square
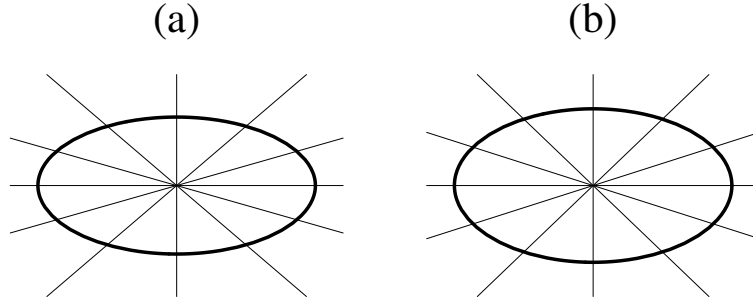
Figure 3.9: Anisotropy ellipses (with phase lines in intervals of $\pi/6$) with the mean eccentricities (a) transverse to dislocation lines in space $\langle \varepsilon \rangle_{d,3}$ from (3.3.9) (b) in the plane $\langle \varepsilon \rangle_{d,2}$ from (3.2.55).

(2.4.5) are identical in two and three dimensions, for ∇, ω defined appropriately. The characteristic speed v_c (3.2.58) is identical for speed in plane sections and in space. The calculation is similar to (3.2.62), and details of the calculation may be found in [BD00]. The resulting probability density function for transverse speed v of a dislocation line is

$$P(v) = \frac{3^{5/2} v v_c^3}{(3v_c^2 + v^2)^{5/2}}. \tag{3.3.10}$$

Only the first two moments do not diverge:

$$\begin{aligned} \langle v \rangle_{d,3} &= \sqrt{3} v_c = c\sqrt{3\left(1 - \frac{k_1^2}{k_2}\right)}, \\ \langle v^2 \rangle_{d,3} &= 6v_c^2 = 6c^2 \left(1 - \frac{k_1^2}{k_2}\right) \end{aligned} \tag{3.3.11}$$

The first moment is less than the first moment of the distribution plane section speed V (3.2.64), as expected from the discussion in section 2.5. The only spectrum with nonzero v_c is the Planck spectrum, with characteristic speed given in (3.2.65). The root mean square transverse speed of dislocation lines in blackbody radiation is therefore $1.146c$.

3.3.4   Statistics of curvature and torsion

The observation (2.7.4) that, in an appropriate reference frame, the squared curvature is equal to the squared speed (with ξ_{zz}, η_{zz} replacing ξ_t, η_t) means that the probability density function for curvature is the same as for speed (3.3.10). This reference frame is chosen when integrating to find the curvature distribution. When this is done (ie averaging the expression (2.7.4), there are only two (identical) nondiagonal correlation matrices (as

with speed), that between ξ and ξ_{zz}, and between η and η_{zz}; they both have the same matrix $\left(\begin{smallmatrix} 1 & -k_2/3 \\ -k_2/3 & k_4/5 \end{smallmatrix}\right)$. This implies, anticipating averaging only on dislocations,

$$P(\xi = 0, \xi_{zz}) = \frac{1}{2\pi\kappa_{\mathrm{c}}\sqrt{k_2}} \exp(-\xi_{zz}^2/2k_2\kappa_{\mathrm{c}}), \qquad (3.3.12)$$

and similarly for η, η_{zz}. κ_{c} is the *characteristic curvature*, defined analogously to characteristic speed v_{c},

$$\kappa_{\mathrm{C}} \equiv \sqrt{\frac{9k_4 - 5k_2^2}{45k_2}}. \qquad (3.3.13)$$

The calculation of the curvature probability density is now identical to the three dimensional speed probability density, so the resulting distribution is

$$P(v) = \frac{3^{5/2}\kappa\kappa_{\mathrm{c}}^3}{(3\kappa_{\mathrm{c}}^2 + \kappa^2)^{5/2}}, \qquad (3.3.14)$$

with first two moments

$$\langle\kappa\rangle_{\mathrm{d}} = \sqrt{3}\kappa_{\mathrm{c}}, \qquad \langle\kappa^2\rangle_{\mathrm{d}} = 6\kappa_{\mathrm{c}}^2. \qquad (3.3.15)$$

Note that the form for κ_{c} is similar to the distribution of phase saddles in plane sections (3.2.17) using (3.1.12), because the averages involve the same correlation matrix (between second derivatives ξ_{ii} and the field ξ). With the shell spectrum, $\kappa_{\mathrm{c}} = 2k_{\mathrm{m}}/\sqrt{45}$, so the mean square curvature is

$$\langle\kappa\rangle_{\mathrm{d}} = 2\langle\kappa\rangle_{\mathrm{d}}^2 = \frac{8k_{\mathrm{m}}^2}{15}; \qquad (3.3.16)$$

a measure of the radius of curvature is

$$1/\sqrt{\langle\kappa\rangle_{\mathrm{d}}} = 0.218\lambda_{\mathrm{m}}, \qquad (3.3.17)$$

indicating that dislocations are sharply curved on the wavelength scale. For the Planck spectrum,

$$\langle\kappa^2\rangle_{\mathrm{d}} = 2\langle\kappa\rangle_{\mathrm{d}}^2 = \frac{5938k_T^2}{1575}. \qquad (3.3.18)$$

The average radius of curvature is approximated by

$$1/\sqrt{\langle\kappa^2\rangle_{\mathrm{d}}} = 0.026\lambda_T. \qquad (3.3.19)$$

If this were measured in units estimated from the peak of the Planck spectrum (3.1.40), this comes to be about 0.2, similar to (3.3.17).

The torsion calculation (using (2.7.6)) is considerably more complicated; not only does it involve third derivatives of ξ, η, but is divided by κ^2, making evaluation of the integrals rather difficult. It has not been possible to calculate the probability density function of torsion, and its second moment seems to diverge.

3.3.5 Statistics of twist and twirl

In this section the probability density functions of the various twist and twirl values from section 2.8 are calculated, and the notation of that section is used, such as $\nabla \xi = \mathbf{X}, \nabla \eta = \mathbf{Y}$. The dislocation may be rotated to be in the z-direction without loss of generality, as in the previous section. We shall calculate the distribution of the phase helicoid twist $Tw(\chi)$ (2.8.2), the azimuth-average phase twist Tw_ϕ (2.8.7), Berry's screwness parameter σ (2.8.8), the phase and azimuth twirls tw_ϕ, tw_χ (2.8.9), (2.8.10), and their difference, the twirl twist Tw_{tw} (2.8.11).

The individual phase helicoid twist $Tw(\chi)$ is easy to evaluate; any phase may be chosen by statistical gauge invariance, and we choose to work with the surface corresponding to $\xi = 0$, which has twist

$$Tw(0) = \frac{\mathbf{T} \cdot \mathbf{X} \wedge \mathbf{X}'}{X^2} \qquad (3.3.20)$$

The only second derivatives appearing are the cross terms ξ_{xz}, ξ_{yz} which are statistically independent of the other functions in the average by (3.1.17). Therefore the probability distribution of $Tw = Tw(0)$ is (by (3.1.24))

$$
\begin{aligned}
P(Tw) &= \langle \delta(Tw - \mathbf{T} \cdot \mathbf{X} \wedge \mathbf{X}'/X^2) \rangle_d \\
&= \frac{3}{4Tw_c} \frac{1}{(1 + Tw^2/Tw_c^2)^{5/2}},
\end{aligned} \qquad (3.3.21)
$$

where details of the calculation are omitted. Tw_c is the *characteristic twist*, defined by

$$Tw_c = \sqrt{\frac{k_4}{5k_2}}. \qquad (3.3.22)$$

The other twists can all be found by using the conditional probability density function $P(f; G, \omega)$, where f is the appropriate twist. For screwness σ, this is

$$P(\sigma; G, \omega) = \sqrt{15/2\pi k_4} \sqrt{G} \exp(-15G\sigma^2/2k_4), \qquad (3.3.23)$$

and the others are the same apart from a different dependence on G, ω. For σ, there is a factor of G in the exponent (multiplied overall by $\sqrt{G}$ for normalisation). This factor is replaced by $4\omega^2/G$ for Tw_ϕ, $(G + 2\omega)$ for Tw_{tw}, and $(G^2 - 3\omega^2)/G$ for the two twirls tw_ϕ, tw_χ. The probability density calculations are similar to that of Tw in (3.3.21), except the distribution (3.1.25) for G and ω is used. The resulting normalised probability density functions of the various twists and twirls are (all are scaled in terms of the characteristic

	First moment $\langle|f|\rangle$	Second moment $\langle f^2\rangle$
Phase helicoid twist Tw	$\frac{1}{2}$	$\frac{1}{2}$
Azimuth-averaged twist Tw_ϕ	$\frac{15}{32}$	$\frac{1}{2}$
Screwness σ	$\frac{5}{16}$	$\frac{1}{2}$
Twirl-averaged twist Tw_{tw}	$\frac{7}{4\sqrt{2}} - 1$	$\frac{1}{2}\left(\log 2 - \frac{1}{2}\right)$
(Phase and ellipse) twirls tw	0.7363	not convergent

Table 3.2: Table of first and second moments of the various twists and twirls, in appropriate units of characteristic twist Tw_{c}, computed from (3.3.21), (3.3.24).

twist Tw_{c} (3.3.22))

$$
\begin{aligned}
P(\sigma) &= \frac{35}{32(1+\sigma^2)^{9/2}}, \\
P(Tw_\phi) &= \frac{3}{32Tw_\phi^4}\left(2 - \frac{(2+7Tw_\phi^2)}{(1+Tw_\phi^2)^{7/2}}\right), \\
P(Tw_{tw}) &= \frac{1}{8}\left(\frac{\sqrt{2}(11+8Tw_{tw}^2+32Tw_{tw}^4)}{(1+2Tw_{tw}^2)^{7/2}} - \frac{4}{(1+Tw_{tw}^2)^{3/2}}\right), \\
P(tw) &= \frac{1}{32tw^4(1+tw^2)^{5/2}}\left((2+tw^2)(3+3tw^2+8tw^4)E\left(\frac{tw^2}{1+tw^2}\right)\right. \\
&\qquad\qquad \left. -2(3+3tw^2+2tw^4)K\left(\frac{tw^2}{1+tw^2}\right)\right), \qquad (3.3.24)
\end{aligned}
$$

where in the last equality, tw represents either tw_ϕ, tw_χ, and K, E represent the complete elliptic integrals of the first and second kinds [AS65]. The five different distributions are plotted in figure (3.10), and their first two moments are given in table (3.2). Although the twists were all equal for the twist example (2.8.3), the distributions are all different for dislocations in random waves.

From the table and figure, it is clear that the distribution of azimuth-averaged twist Tw_ϕ is closest to the twist of a single phase surface $Tw = Tw(\chi)$.

The characteristic twist of the shell spectrum is $Tw_{\mathrm{c}} = k_{\mathrm{m}}/\sqrt{5}$. A measure of the mean pitch of a single phase helicoid is

$$
2\pi/\sqrt{\langle Tw^2\rangle_{\mathrm{d}}} = \sqrt{10}\lambda_{\mathrm{m}}. \qquad (3.3.25)
$$
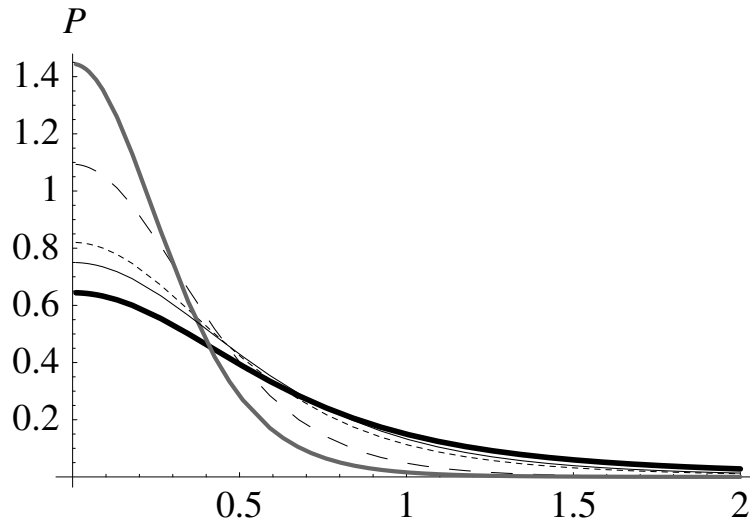
Figure 3.10: The probabilities of the different twists and twirls, given by equations (3.3.21), (3.3.24). In ascending order up the y-axis, they are: twirl tw (thick black line), helicoid twist Tw (thin black line), azimuth-averaged twist Tw_ϕ (dotted line), screwness σ (dashed line), twirl-averaged twist Tw_{tw} (thick grey line).

3.4 Discussion and Conclusions

In this chapter, we have used a natural isotropic random wave model to calculate the statistics of geometric quantities of dislocation points in three dimensions, and dislocation lines in three. Apart from the calculations involving speed, the wave equation played only a minor role, and, with appropriate modifications, many of the averages apply to ensembles of more general random functions.

For dislocation points in the plane, the interesting questions that remain are connected with the various correlation functions, and were discussed in detail in section 3.2.3. As indicated there, it is possible that the ring spectrum correlation functions can be used to explain the morphological scarlet structures observed by [OGH87], although so far, investigation of this question has not produced any interesting new results. Further insight may be gained from a calculation of the critical point correlation functions (with Poincaré index playing the role of charge). However, the multiple integral necessary to evaluate these functions has so far proved intractable.

For dislocation lines, of the quantities averaged here (core ellipse eccentricity, speed, curvature and the various phase twists) have been *local* properties, depending only on

the local derivatives at a point on the dislocation (detailed torsion statistics were not possible, and the distribution of Tw_χ (2.8.2) was not even attempted). More interesting statistics are global (that is, topological). Are dislocation lines, on average, infinitely long or closed loops? We shall see in chapter 5 that dislocation loops may be knotted, but does knottedness (or linking) occur with finite probability? If so, what is the distribution of knot invariants (screw numbers), etc? It is not clear how one should even go about trying to evaluate these global statistics.

Chapter 4

Polarization singularities in vector waves

> ' "In giving to these sides [of the vertical ray] the name of *poles*, [Malus]
> calls the modification which imparts to light properties relative to these poles,
> polarization ... " But this unfortunately assumed a sense of pole quite different
> from its use in astronomy, geography, and magnetism, with the consequence
> that polarization as applied to light and radiant heat has nothing in common
> with magnetic or electric polarization.'

From the etymology for POLARIZE in *the Oxford English Dictionary*, 2nd edition, 1989
(the quotation is translated from *Nouveau Bulletin de Sciences* **45**, March 1811)

This chapter is concerned with the details of polarization singularities, their nature, and their statistical densities in random vector waves in both two and three dimensions. General results concerning ellipses are in appendix A, and the statistical part relies heavily on the previous chapter (particularly section 3.1). Some of the earlier material in this chapter is well-known; some of the new results and statistics for polarization singularities in three dimensions have been published in [BD01c]. The structures described here are generalised to singularities in tensor waves in chapter 6.

4.1   Polarization: The vector nature of vector waves

The waves considered in chapters 2, 3 were scalar fields, in which the topological singularities were nodal points and lines. However, many waves in physics, particularly electromagnetic radiation and light, are physically described by vectors (or, more generally, tensors), requiring additional physical parameters other than phase and amplitude to describe them. This additional structure shall be referred to as the *polarization* structure of the field, and topological singularities (of codimension 1,2,3) in such waves are singularities of the polarization structure rather than phase.

As the author of the quotation above appreciated, the 'poles' of a polarized plane wave are rather different from the poles of a magnet; the polarization ellipse is invariant with respect to a rotation by π, unlike a magnet (which also requires the poles to change sign). This rotational symmetry, as we saw in chapter 1, implies that the index of a point of circular polarization in a vector wavefield is in units of 1/2, rather than 1.

As with scalar waves, the vector wavefields we shall use are complex, where, if necessary, the complex analytic signal of a real vector field is used. The topological singularities in such fields are therefore different from the saddles, sources, sinks and circulations of real vector fields. As phase is varied, the real part of the complex vector traces out an ellipse (the *polarization ellipse*), and the polarization singularities occur when the ellipse is circular (C singularities) and when it is linear (L singularities). These singularities are invariant with respect to a global phase change, as were the dislocations studied earlier. In paraxial vector waves, all of the component plane waves propagate in (almost) the same direction, so the vectors are confined to the same transverse plane, and the polarization ellipse field resides in this plane [Nye83b, Nye83a]. Although these references consider the paraxial field in three dimensions, we shall only consider the field in the transverse plane, and so have a two-dimensional vector field in two dimensions. In the more general nonparaxial case, in three dimensions, the plane wave components may be propagating in any direction, and the planes of the polarization ellipses at different places in space are different [NH87].

The codimension of L singularities, unlike C singularities or dislocations, is sensitive to the dimension of the wavefield (ie whether it is paraxial or not), as we shall see. More care must therefore be taken than for scalar waves as to the nature of the field, and we shall treat the two cases separately (there are, of course, similarities between the two

cases as well). Although much of the discussion applies to any (complex) vector field, we often shall be concerned when the field is the electric field describing light. In section 4.4, possible singularities involving the entire electromagnetic field (involving both electric and magnetic fields) are discussed.

We begin with a discussion of two-dimensional polarization singularities, since these are easier to visualise and study experimentally, have a simpler geometry, and were historically the first to be found.

4.2   Polarization singularities in two dimensions

4.2.1   The polarization ellipse

We consider a two dimensional vector field $\mathbf{E} : \mathbb{R}^2 \longrightarrow \mathbb{C}^2$, possibly time-dependent, and the complex vector $\mathbf{E} = \mathbf{E}(\mathbf{R}, t)$ is written in terms of its real and imaginary parts

$$\mathbf{E} = \mathbf{P} + i\mathbf{Q}. \tag{4.2.1}$$

As in equation (A.2.3), the real vector $\operatorname{Re} \mathbf{E} \exp(-i\chi)$ traces out an ellipse as the real phase χ varies. As χ increases, the ellipse is traced out either anticlockwise (right handed, or positive polarization), or clockwise (left handed, or negative polarization). These conventions differ from that of some authors. The field of ellipses is therefore invariant under any phase transformation $\mathbf{E} \to \mathbf{E} \exp(i\chi)$, as are the Stokes parameters (A.4.7) describing the ellipse at each point. For monochromatic fields, time acts like the phase; the physical disturbance is the real part of $\mathbf{E}$ as phase varies, and the entire ellipse field is stationary. We shall not consider the possibility that $\mathbf{E}$ is partially polarized, and time dependence shall not play an important role in the following discussion.

The squared modulus of the vector $\mathbf{E}$, $|\mathbf{E}|^2 = |\mathbf{P}|^2 + |\mathbf{Q}|^2$, plays the role of intensity for vector waves, and shall be referred to as the *intensity* of the ellipse. It corresponds to the Stokes parameter S_0 (equation (A.4.7)). It only vanishes when both vectors $\mathbf{P}, \mathbf{Q}$ are zero, which occurs with codimension 4, so does not happen generically.

For complex vector $\mathbf{E}$, the rectifying phase angle χ_0 (defined modulo π), is defined:

$$\tan 2\chi_0 = \frac{2\mathbf{P} \cdot \mathbf{Q}}{P^2 - Q^2}, \tag{4.2.2}$$

(see also equation (A.2.6) and the discussion there); the real and imaginary parts $\mathbf{P}_0, \mathbf{Q}_0$ of $\mathbf{E} \exp(-i\chi_0) = \mathbf{P}_0 + i\mathbf{Q}_0$ are orthogonal. This angle was called *phase of the vibration* by

[Nye83a]. The smallest value of χ_0 in the arctangent is taken, meaning either $\mathbf{P}_0$ or $\mathbf{Q}_0$ can be the major semiaxis, but we shall usually assume it to be $\mathbf{P}_0$. The angle the major semiaxis $\mathbf{P}_0$ makes with the x-axis is denoted by γ, so

$$\tan\gamma = \frac{P_{0y}}{P_{0x}}. \tag{4.2.3}$$

The ellipse can also be described by coordinates on the Poincaré sphere, (that is, by normalised Stokes parameters s_1, s_2, s_3), with azimuthal angle $\beta = 2\gamma$. We also define the phase-invariant quantity

$$\begin{aligned} N &= \frac{1}{2}\mathrm{Im}(\mathbf{E}^* \wedge \mathbf{E}) \\ &= \mathbf{P} \wedge \mathbf{Q} \\ &= \mathbf{P}_0 \wedge \mathbf{Q}_0 \\ &= P_x Q_y - P_y Q_x, \end{aligned} \tag{4.2.4}$$

which is the $\mathbf{E}$-analogue of ω defined for the complex vector $\nabla\psi$ in equation (2.2.2). As with ω, the sign of N defines the handedness of the ellipse ($N > 0$ right handed, $N < 0$ left handed), and $2N$ is the Stokes parameter S_3, proportional to the area of the ellipse. From (A.4.9), the ellipse eccentricity ε can be found from

$$\varepsilon^2 = \frac{1}{2N^2}\sqrt{|\mathbf{E}|^4 - 4N^2}\left(|\mathbf{E}|^2 - \sqrt{|\mathbf{E}|^4 - 4N^2}\right). \tag{4.2.5}$$

As with ω defined in the plane, N may sometimes be considered as a vector in the z-direction. There are four real parameters P_x, P_y, Q_x, Q_y describing the field at each point; four geometric parameters describing the polarization ellipse are size of the ellipse (intensity $|\mathbf{E}|^2$), phase of the ellipse (rectifying phase χ_0), ellipse orientation angle γ, and eccentricity ε, as discussed in section 1.3. There is, in addition, one discrete parameter, the handedness of the polarization. At C points, χ_0 and γ are singular, and on L lines, the ellipse handedness is singular.

4.2.2   C points

The codimension of a C locus is 2 (ie a point in the plane), since there are two conditions for the ellipse traced out by $\mathrm{Re}\,\mathbf{E}\exp(-i\chi)$ to be a circle [Nye83a, NH87, BD01c]: $\mathbf{P}, \mathbf{Q}$ must be equal in length (since they are radii of the circle) and orthogonal (since they are conjugate semidiameters of the ellipse, as shown in equation (A.2.7)). Defining the

polarization scalar φ by

$$\varphi = u + \mathrm{i}v \equiv \mathbf{E} \cdot \mathbf{E} = P^2 - Q^2 + 2\mathrm{i}\mathbf{P} \cdot \mathbf{Q}, \tag{4.2.6}$$

the C conditions are satisfied if

$$\varphi = 0 \qquad \text{at a C point.} \tag{4.2.7}$$

Complex vectors with a vanishing scalar product with themselves are called *null* [Syn58, PR84b] or *isotropic* [Car66], (not to be confused with the use of the word isotropic employed for wave superpositions), and have important and interesting mathematical properties (which generalise to waves of higher spin, as explored in chapter 6).

There is no unique set of orthogonal semiaxes for a circle (any orthogonal pair of radii suffice), so a C point is a singularity of the rectifying phase χ_0, and indeed, this is the phase which is singular, since from (4.2.2),

$$\arg \varphi = 2\chi_0. \tag{4.2.8}$$

Also,

$$\begin{aligned}
|\varphi|^2 &= (P^2 - Q^2)^2 + 4(\mathbf{P} \cdot \mathbf{Q})^2 \\
&= P^4 + Q^4 - 2P^2Q^2 + 4(\mathbf{P} \cdot \mathbf{Q})^2 \\
&= (P^2 + Q^2) - (2\mathbf{P} \wedge \mathbf{Q})^2 \\
&= S_0^2 - S_3^2 = S_0^2(1 - s_3^2) \\
&= P_0^2 \varepsilon^2,
\end{aligned} \tag{4.2.9}$$

so the interpretation of the modulus of φ is the length of the major semiaxis times the eccentricity of the ellipse, which is zero for a circle.

Geometrically, the fact that a unique semiaxis frame $(\mathbf{P}_0, \mathbf{Q}_0)$ cannot be defined at a C point implies that the orientation angle $\gamma = \beta/2$ is singular, suggesting an alternative polarization scalar φ_S (the *Stokes scalar*), with zeros the C points, defined by

$$\varphi_S \equiv S_1 + \mathrm{i}S_2 = \sqrt{S_0^2 - S_3^2} \exp \mathrm{i}\beta. \tag{4.2.10}$$

This has the same modulus as φ by (4.2.9), but phase β instead of $2\chi_0$. These two scalar fields, φ and φ_S, are not equivalent, although they have the same zeros (since they have the same modulus). φ is rotation invariant, but is dependent on the overall phase. φ_S, on

the other hand, only depends on the ellipse geometry, and is completely phase invariant. However, any rotation of the coordinate system changes the phase of φ_S by twice that angle (since $\beta = 2\gamma$).

There is no simple geometric transformation between the two scalars, and they contain different information, φ about the ellipse phase, φ_S the ellipse geometry. The $u = 0$ contours of φ are the locus of places where the conjugate semiaxes $\mathbf{P}, \mathbf{Q}$ are of equal length, $v = 0$ where they are orthogonal; $S_1 = 0$ occurs when the ellipse semiaxes are in the directions of the lines $x = 0, y = 0$, $S_2 = 0$ when they are parallel to $x = y, x = -y$. The existence of these two scalars is accounted for by the fact that two different geometric quantities, γ and χ_0, are singular at a C point.

The complex vector $\mathbf{E}$ can be written in a circular basis, in which it has components α_+, α_-, as defined and discussed in section A.3, one of which is zero at a C point ($\alpha_- = 0$ for a right handed C point and vice versa). This provides a third way of identifying C points: as dislocations in the scalar fields α_+, α_-, a zero in $\alpha_\pm$ being a C point with the opposite handedness $\mp$.

As described in chapter 1, the generic C points have index $\pm 1/2$ ($+1/2$ for lemon and monstar, $-1/2$ for star), the sign corresponding to the sense in which γ rotates in a circuit around the C point. Therefore, the sign of the (index ± 1) phase singularity in the Stokes scalar φ_S gives the correct sign of the C point index, the factor 2 from the fact that a change of 2π in β implies γ changes by π.

Berry [Ber01a] constructed a simple field with a right-handed C point at the origin,

$$\mathbf{E}(x, y) = (1 + \mathrm{i}ay, \mathrm{i}(1 - x)), \qquad (4.2.11)$$

which is a has index $-1/2$ (star) if $a < 0$, and index $+1/2$ if $a > 0$ (lemon if $a < 2$, monstar otherwise). These fields, with the vectors $\mathbf{P}, \mathbf{Q}$, are shown in figures (4.1) (lemon) and (4.2) (star), for $a = \pm 1$.

The relationship between the sign of a phase singularity in the polarization scalar φ and the C point index is more complicated than for the Stokes scalar. For a right-handed C point, the sign of the phase singularity in φ gives the correct sign of the C point. This is because in the neighbourhood of the C point, neither of $\mathbf{P}, \mathbf{Q}$ are generically zero, so the $(\mathbf{P}, \mathbf{Q})$ frame, although not orthogonal, is smoothly defined (neither vector has any circulations, nor are they collinear). Therefore the ellipses rotate about the $(\mathbf{P}, \mathbf{Q})$ phase frame as well as the coordinate frame, giving the same index as the rotation. However, a
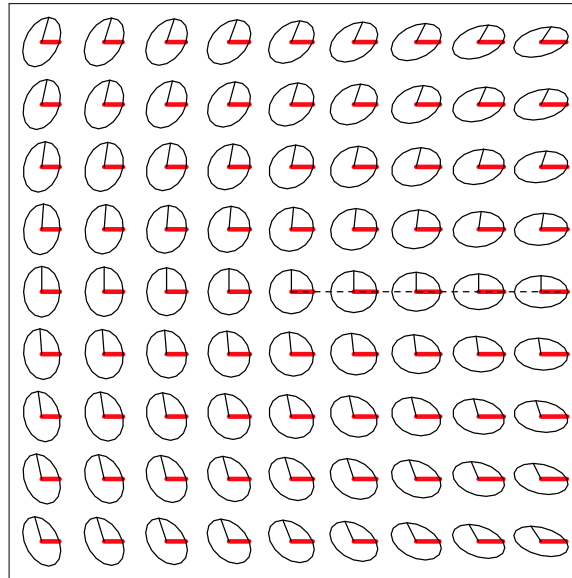
Figure 4.1: A C point with index $+1/2$ (lemon) from (4.2.11) with $a = 1$. The thicker line is the real part $\mathbf{P} = (1, 0)$, (see also figure (1.5)).
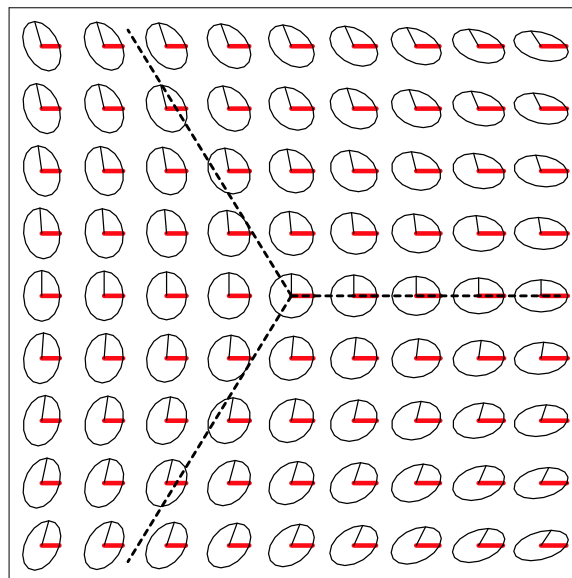
Figure 4.2: A C point with index $-1/2$ (star), from (4.2.11) with $a = -1$. The thicker line is the real part $\mathbf{P} = (1, 0)$.

left-handed C point behaves like the complex conjugate of a right-handed C point, so χ_0 decreases around a left-handed C point where it increases around a right-handed C point. This implies that, around an index $+1/2$ C point, zeros of φ have strength -1, and vice versa.

This can be understood more clearly with a simple example. Define the field

$$\mathbf{E}_R = (c + x, y) + \mathrm{i}(x, b + y), \tag{4.2.12}$$

where c is a real positive constant, and b is real. Here, $N = c^2 + c(x + y)$, so the C point is right-handed. For this field, up to linear terms only,

$$\varphi = \varphi_S = 2c((x - y) + \mathrm{i}(x + y)) \tag{4.2.13}$$

which is easily verified to have a $+1$ phase singularity at the origin, and is of lemon type. By comparison, the function

$$\mathbf{E}_L = (c + x, y) - \mathrm{i}(x, b + y), \tag{4.2.14}$$

has a left-handed lemon C point at the origin, with polarization scalars (including linear terms only)

$$\varphi = \varphi_S^* = 2c((x - y) - \mathrm{i}(x + y)) \tag{4.2.15}$$

which has topological charge -1 at the origin. The Stokes parameters S_1, S_2 are the same for each of $\mathbf{E}_R, \mathbf{E}_L$.

This example may also be used to see what happens with the sign of the index of the C point and the dislocation strengths of the zeros of the circular components α_+, α_- : the sign of the dislocation in α_- corresponding to a right-handed C point is the same as that of the C point index, whereas the sign of a dislocation in α_+ corresponding a left-handed C point is minus the sign of the C point index.

C points occur in the Poincaré sphere description at the poles, (right-handed at the north pole, left at the south, the index of each being $+1/2$), and many of the points made above correspond to a geometrical argument on the Poincaré sphere. The alternative polarization scalars φ, φ_S also imply that there is another geometrical object analogous to the Poincaré sphere: this *phase-Poincaré sphere* has the same colatitude as the corresponding Poincaré sphere (ie same third cartesian component $s_3 = 2N/|\mathbf{E}|^2$), but with azimuth $2\chi_0$ instead of β, implying that the *phase-Stokes parameters* are defined

$$(T_0, T_1, T_2, T_3) = (S_0, \operatorname{Re}\varphi, \operatorname{Im}\varphi, S_3), \tag{4.2.16}$$

which shall be used in chapter 6.

Since C points have been realised as phase singularities, the mathematics of chapter 2 applies, although φ, φ_S are not solutions to any wave equation, and are quadratic in the field variables. There are two sign rules, one for each of the scalars; it seems that C points of different handedness cannot generically be created/annihilated in pairs, since opposite index, oppositely handed C points have the same strength in φ. Regions of right-handed polarization and left-handed polarization are separated by the second type of paraxial polarization singularity, the L line.

4.2.3   L lines and disclinations

The polarization ellipse becomes a line when the angle between $\mathbf{P}, \mathbf{Q}$ vanishes (modulo π), and the real vector $\operatorname{Re} \mathbf{E} \exp(-\mathrm{i}\chi)$ points in the same direction for all χ. This occurs when

$$N = \mathbf{P} \wedge \mathbf{Q} = S_3/2 = 0. \tag{4.2.17}$$

This requires the satisfaction of a single equation, implying that the polarization is linear along lines; L singularities, in two dimensions, have codimension 1 ([Nye83a] calls them S lines). Mathematically, on L lines, the complex vector $\mathbf{E}$ becomes a real vector times a phase, and the singularity in this case is the handedness of the ellipse. (Codimension 1 singularities are usually singularities of a discrete variable, whereas higher codimension implies singularity of a continuous variable [Mer79]). Generically, the L lines separate regions of the plane of different handedness.

On a circuit around a closed L line, the change in the line orientation γ divided by 2π corresponds to the total index of C points enclosed by the L line. For instance, the equator of the Poincaré sphere is a closed L line, where the linear ellipses rotate by $+\pi/2$ as β increases from 0 to 2π; there is a single lemon-type C point at the north pole.

The direction of the L line at a point can be found by finding the direction in which N remains zero; if the point is taken to be the origin, the direction of the L line is $\mathbf{R}$, where

$$\mathbf{R} \cdot \nabla N = 0, \tag{4.2.18}$$

ie a vector orthogonal to ∇N. The direction $\mathbf{D}_\mathrm{L}$ is therefore

$$\mathbf{D}_\mathrm{L} = (P_x Q_{y,y} - P_y Q_{x,y} + P_{x,y} Q_y - P_{y,y} Q_x, -P_x Q_{y,x} + P_y Q_{x,x} - P_{x,x} Q_y + P_{y,x} Q_x), \tag{4.2.19}$$

and along the direction of $\mathbf{D}_\mathrm{L}$, the region with right handed polarization is to the left. To find the total length $\ell_{\mathrm{L}\ \mathcal{A}}$ of L line in an area $\mathcal{A}$ of the plane, equations (2.1.6), (2.5.1) are appropriately modified,

$$\ell_{\mathrm{L}\ \mathcal{A}} = \int_{\mathcal{A}} \mathrm{d}^2\mathbf{R}\, \delta(N)|\nabla N|. \tag{4.2.20}$$

We observe that $|\mathbf{D}_\mathrm{L}| = |\nabla N|$.

There is another type of singularity, not phase invariant, called a wave disclination, introduced by Nye [Nye83b]. These are places where, for a given phase χ, the real vector

$$\mathbf{U}_\chi(\mathbf{R}) \equiv \operatorname{Re}\mathbf{E}(\mathbf{R}) \exp(-\mathrm{i}\chi) \tag{4.2.21}$$

vanishes. If $\chi = 0\ (\pi/2)$, these are the zeros of $\mathbf{P}$ ($\mathbf{Q}$). Clearly, they have codimension 2, and are always found along L lines. At any point on an L line, there is always a phase for which that point is a disclination.

As phase varies (say with time), the disclinations move along the L line periodically with period π. Not being invariants of phase, we shall not consider them in any more detail, but notice that the number of disclinations in $\mathcal{A}$ for phase χ is

$$d_{\mathrm{disc}\ \mathcal{A}} = \int_{\mathcal{A}} \mathrm{d}^2\mathbf{R}\, \delta^2(\mathbf{U}_\chi)|U_{\chi x,x}U_{\chi y,y} - U_{\chi x,y}U_{\chi y,x}|. \tag{4.2.22}$$

A disclination is a zero of the field $\mathbf{U}_\chi$, and generically has Poincaré index ± 1 [Nye83b]. Of course, if one considers three-dimensional paraxial fields, disclinations are lines, as described by [Nye83b].

4.3 Polarization singularities in three dimensions

4.3.1 The polarization ellipse in three dimensions

We are now interested in complex solutions of a three-dimensional vector wave equation $\mathbf{E}: \mathbb{R}^3 \longrightarrow \mathbb{C}^3$, where

$$\mathbf{E}(\mathbf{r}) = \mathbf{p}(\mathbf{r}) + \mathrm{i}\mathbf{q}(\mathbf{r}), \tag{4.3.1}$$

and $\mathbf{p}, \mathbf{q}$ are three-dimensional vectors. As phase varies, the real vector

$$\mathbf{f}_\chi = \operatorname{Re}\mathbf{E}\exp(-\mathrm{i}\chi). \tag{4.3.2}$$

traces out an ellipse in the $(\mathbf{p}, \mathbf{q})$ plane. The rectifying phase χ_0 for the ellipse in three dimensions is defined analogously to (4.2.2):

$$\mathbf{E}\exp(-\mathrm{i}\chi_0) = \mathbf{p}_0 + \mathrm{i}\mathbf{q}_0, \tag{4.3.3}$$

with $\mathbf{p}_0, \mathbf{q}_0$ perpendicular. The normal to the ellipse is given by

$$\mathbf{N} = N\mathbf{n} \equiv \mathbf{p} \wedge \mathbf{q} = \mathbf{p}_0 \wedge \mathbf{q}_0, \tag{4.3.4}$$

where N here corresponds to the modulus of the definition for N in two dimensions (4.2.4). As before, and with ω defined previously, N is proportional to the area of the ellipse. The ellipse is always right-handed with respect to the unit vector $\mathbf{n}$. The field at a point is determined by six real numbers (the components of $\mathbf{p}, \mathbf{q}$); these can be interpreted geometrically as the ellipse intensity $|\mathbf{E}|^2$, the rectifying phase χ_0, the eccentricity ε, but the orientation of the ellipse in space now has three parameters: the direction of $\mathbf{n}$ providing two, and the angle the ellipse makes in the plane normal to $\mathbf{n}$, which as before shall be called γ (these three angles, γ and the polar angles defining $\mathbf{n}$, are equivalent to the three Euler angles). Note that there is no smooth canonical way of ascribing a direction for $\gamma = 0$ for all $\mathbf{n}$. The handedness of the ellipse no longer has any meaning. The ellipse gives rise to a positively oriented orthogonal frame field $(\mathbf{p}_0, \mathbf{q}_0, \mathbf{n})$, which are the eigendirections of the matrix

$$\mathbf{M} = \operatorname{Re} \mathbf{E}^* \otimes \mathbf{E} = \mathbf{p} \otimes \mathbf{p} + \mathbf{q} \otimes \mathbf{q}, \tag{4.3.5}$$

as discussed in section A.2. It is this frame that becomes singular at polarization singularities [NH87, BD01c]. Note that the C lines and L lines defined here are called C^T, L^T lines by [NH87].

4.3.2 C lines in three dimensions

The C conditions in three dimensions are analogous to those in two, and the polarization scalar φ can be similarly defined:

$$\varphi = u + \mathrm{i}v = \mathbf{E} \cdot \mathbf{E} = p^2 - q^2 + 2\mathrm{i}\mathbf{p} \cdot \mathbf{q}. \tag{4.3.6}$$

The phase singularities of φ are C lines, where $u = 0, v = 0$. As with equations (4.2.8), (4.2.9), $\arg \varphi = 2\chi_0$, $|\varphi| = p_0 \varepsilon$, and $\mathbf{p}_0, \mathbf{q}_0$ do not uniquely exist at a C point as before. The matrix $\mathbf{M}$ (4.3.5) is degenerate in its nonzero eigenvalues in this case. The degeneracies of real symmetric matrices are well-known to have codimension 2, and are called diabolical points (see, for instance, [Ber84] page 50).

The Poincaré sphere is not defined for three-dimensional polarization, and φ_S cannot be defined. The Majorana sphere $\mathcal{M}_2$ can be used instead, and the details of this are

explained in chapter 6. The problem is due to the lack of any canonical way of defining γ for every ellipse orientation in three dimensions. The phase-Poincaré sphere, however, is defined for complex vectors in three dimensions.

As with the three-dimensional definition of vorticity $\boldsymbol{\omega}$ in section 2.5,

$$\boldsymbol{\omega}_\varphi = \frac{1}{2} \nabla \varphi^* \wedge \nabla \varphi = \nabla u \wedge \nabla v \qquad (4.3.7)$$

points in the direction of the C line, giving a form for the topological current. The direction of the ellipse normal vector $\mathbf{n}$ is unrelated to the direction of the C line. However, in the vicinity of a point $\mathbf{r}_0$ on the C line, the local ellipses are confined to the plane normal to $\mathbf{n}(\mathbf{r}_0)$ and the C point in this plane generically has index $1/2$, with lemon, star or monstar morphology. The sign of the index of this C point is the sign of $\boldsymbol{\omega}_\varphi \cdot \mathbf{n}$, as described by [NH87] (see in particular figure 2). This is consistent with the relation between index of C points and phase singularities of φ, described in section 4.2.2, where the sign of the C point index was seen to be the product of the handedness of the circle with the topological charge of φ; $\mathbf{n}$ points in the direction with respect to which the ellipse is right handed. Along a C line, its the index switches at points where $\mathbf{n} \cdot \boldsymbol{\omega}_\varphi = 0$. The topological current of the C line, however, does not change.

Any geometric property of dislocation lines (curvature, torsion, anisotropy ellipse) can be found for C lines using the polarization scalar φ, although we do not provide a geometric interpretation of the anisotropy ellipse, or the twist and twirl of a C line.

4.3.3   L lines and disclinations

The L condition for three-dimensional vectors is that $\mathbf{p}, \mathbf{q}$ should be (anti)parallel, which now is two conditions (the direction (two polar angles) of $\mathbf{q}$ corresponds to $\mathbf{p}$). L singularities therefore are lines in three dimensions, and are places where the ellipse normal $\mathbf{N}$ is zero:

$$\mathbf{N} = \mathbf{p} \wedge \mathbf{q} = 0 \qquad \text{on an L line.} \qquad (4.3.8)$$

The codimension of this phenomenon is only 2, although it is a zero of a three-dimensional vector field; $\mathbf{N}$ is a cross product, and vanishes when $\mathbf{p}, \mathbf{q}$ are parallel, which only requires equality of the two sets of polar angles of $\mathbf{p}, \mathbf{q}$. Three-dimensional L lines are singularities of the direction of the ellipse normal $\mathbf{n}$, where the intermediate eigenvalue of the matrix $\mathbf{M}$ (4.3.5) becomes zero, and the matrix is degenerate again. The direction $\mathbf{D}_L$ of the L

line, to be found below, is unrelated to the direction in which $\mathbf{p}, \mathbf{q}$ and $\mathbf{f}_\chi$ coincide; we shall denote the unit vector in the direction of $\mathbf{p}$ by $\mathbf{e_f}$.

Local to a point on the L line, the nearby $\mathbf{n}$ vectors are all coplanar, in the plane normal to the direction of $\mathbf{e_f}$. It is therefore a singularity in a (locally) two-dimensional vector field, and can be a source, sink, circulation or saddle. The index of the L line switches when $\mathbf{D}_\mathrm{L} \cdot \mathbf{e_f} = 0$, as described by [NH87].

We now derive a general expression for the direction $\mathbf{D}_\mathrm{L}$ of the L line, taken initially to be crossing the origin with $\mathbf{e_f}$ in the z-direction. In this case, we have the local expansion (where $\mathbf{r} = (x, y, z)$)

$$
\begin{aligned}
\mathbf{p}(\mathbf{r}) &= (\mathbf{r} \cdot \nabla p_x, \mathbf{r} \cdot \nabla p_y, p_z), \\
\mathbf{q}(\mathbf{r}) &= (\mathbf{r} \cdot \nabla q_x, \mathbf{r} \cdot \nabla q_y, q_z),
\end{aligned}
\tag{4.3.9}
$$

so, to first order

$$
\begin{aligned}
\mathbf{N}(\mathbf{r}) &= (\mathbf{r} \cdot (q_z \nabla p_y - p_z \nabla q_y), \mathbf{r} \cdot (p_z \nabla q_x - q_z \nabla p_x), 0) \\
&= (\mathbf{A} \cdot \mathbf{r}, \mathbf{B} \cdot \mathbf{r}, 0).
\end{aligned}
\tag{4.3.10}
$$

and if $\mathbf{r}$ lies in the direction of the L line, $\mathbf{N}(\mathbf{r}) = 0$ in (4.3.10). This direction must therefore be in the direction of

$$
\begin{aligned}
\mathbf{D}_\mathrm{L} &= \mathbf{A} \wedge \mathbf{B} \\
&= (q_z \nabla p_y - p_z \nabla q_y) \wedge (p_z \nabla q_x - q_z \nabla p_x) \\
&= \nabla(\mathbf{p} \wedge \mathbf{q})_x \wedge \nabla(\mathbf{p} \wedge \mathbf{q})_y.
\end{aligned}
\tag{4.3.11}
$$

This can be written in coordinate-free form as

$$
\mathbf{D}_\mathrm{L} = \frac{1}{2} \nabla_a \wedge \nabla_b (\mathbf{N}_a \wedge \mathbf{N}_b \cdot \mathbf{e_f}),
\tag{4.3.12}
$$

where a, b are labels showing where the ∇ operators act (this is a notation suggested by Feynman [FLS63a]). (4.3.10) can be used to find the necessary behaviour in $\mathbf{E}$ to make the L line have the various vector field singularity morphologies described above. The choice of $\mathbf{e_f}$ (rather than $-\mathbf{e_f}$) was arbitrary, and making this change reverses the direction of $\mathbf{D}_\mathrm{L}$; the sign of the product $\mathbf{D}_\mathrm{L} \cdot \mathbf{e_f}$, remains the same, and the singularity does not change index.

We now wish derive an expression for the total length of L line in a volume $\mathcal{V}$, again choosing coordinates such that the L line crosses the origin and $\mathbf{e_f}(\mathbf{0})$ is in the z-direction

(so $\mathbf{D}_{\mathrm{L}} = \mathbf{A} \wedge \mathbf{B}$). The correct δ-functions, restricting the integral to the L line, are $\delta(\mathbf{A} \cdot \mathbf{r})$ in x, $\delta(\mathbf{B} \cdot \mathbf{r})$ in y, from (4.3.10), whose transverse jacobian is

$$\partial_x \mathbf{A} \cdot \mathbf{r}\, \partial_y \mathbf{B} \cdot \mathbf{r} - \partial_y \mathbf{A} \cdot \mathbf{r}\, \partial_x \mathbf{B} \cdot \mathbf{r} = (\mathbf{A} \wedge \mathbf{B})_z \tag{4.3.13}$$

divided by the cosine of the angle between $\mathbf{D}_{\mathrm{L}}$ and $\mathbf{e_f}$. This cosine is $(\mathbf{A} \wedge \mathbf{B})_z / |\mathbf{A} \wedge \mathbf{B}|$, giving a net jacobian factor of $|\mathbf{A} \wedge \mathbf{B}| = |\mathbf{D}_{\mathrm{L}}|$. Moreover, using the δ-function identity

$$\delta(X)\delta(Y) = \frac{\delta(\sqrt{X^2 + Y^2})}{\pi\sqrt{X^2 + Y^2}}, \tag{4.3.14}$$

and (4.3.10), the length of L line $\ell_{\mathrm{L}\,\mathcal{V}}$ in a volume $\mathcal{V}$ is

$$\ell_{\mathrm{L}\,\mathcal{V}} = \int_{\mathcal{V}} \mathrm{d}^3\mathbf{r}\, \frac{\delta(N)}{\pi N} |\mathbf{D}_{\mathrm{L}}|. \tag{4.3.15}$$

As with (2.5.1) and (4.2.20), the length of the simplest vector in the direction of the singular line homogeneous in the field variables gives the correct jacobian.

Disclinations are defined for spatial waves in the same way as for paraxial waves (zeros of $\mathbf{f}_\chi$ for phase χ) and are points on the L lines. The number of disclination points in volume $\mathcal{V}$ is simply

$$d_{\mathrm{disc}\,\mathcal{V}} = \int_{\mathcal{V}} \mathrm{d}^3\mathbf{r}\, \delta^3(\mathbf{f}_\chi) |\nabla \mathbf{f}_\chi|. \tag{4.3.16}$$

C lines and L lines can only cross if the ellipse at that point is both circular and linear, that is, it vanishes (and $\mathbf{E} = \mathbf{0}$). This phenomenon has codimension 6 (all of the six field variables are zero), which is more of a restriction than the codimension 4 crossing of C lines with themselves or L lines with themselves. Therefore, C lines and L lines repel, and their points of intersection the true phase singularities of the vector field, where the amplitude of the vector field is zero and phase is undefined.

4.4   Polarization singularities in electromagnetic waves

The previous section described topological singularities arising out of the structure of complex three-dimensional vector wavefields. However, the full theory of electromagnetism (in free space) involved two such fields, the electric field $\mathbf{E}$ and also the magnetic field $\mathbf{H}$, both of which are related to the magnetic vector potential $\mathbf{A}$ (in appropriate units):

$$\mathbf{E} = -\partial_t \mathbf{A}, \qquad\qquad \mathbf{H} = \nabla \wedge \mathbf{A}. \tag{4.4.1}$$

All three fields in free space are solutions of the same vector wave equation (1.5.1), and it is a natural physical question to ask whether there is a single singularity structure involving the entire electromagnetic field.

Since $\mathbf{E}$ and $\mathbf{H}$ are derived from $\mathbf{A}$ by (4.4.1), we start by examining the polarization singularities of the vector potential $\mathbf{A}$. There is immediately a problem, because the positions of the polarization singularities are gauge dependent; the transformation $\mathbf{A} \to \mathbf{A} + \mathbf{A}_0$, where $\mathbf{A}_0$ is even a constant vector field, ruins the delicate polarization structure of a singularity. A natural gauge to choose is the one (explicitly taken in (4.4.1)), for which the scalar potential vanishes, as is done by [BW59], page 73. This requires transverseness of $\mathbf{A}$, $\nabla \cdot \mathbf{A} = 0$.

If the field in this gauge is monochromatic, then time dependence may be factored out, and the C and L lines are in the same places in the $\mathbf{E}$ and $\mathbf{A}$ fields. However, there is no clear connection between the singularities of $\mathbf{A}$ and its curl, $\mathbf{H}$, which involves the differential structure of the field in a nontrivial way. There does not therefore seem to be a connection between the polarization singularities in $\mathbf{E}$ and $\mathbf{H}$, even in the field is monochromatic. We shall see in section 4.6 that in the model of isotropic spatial distributions of random plane electromagnetic waves, the components of $\mathbf{E}$ and $\mathbf{H}$ at a point are statistically independent.

Many authors of more abstract texts [LL75, Syn58, PR84b] use the complex vector

$$\mathbf{V} = \mathrm{Re}\,\mathbf{E} + \mathrm{i}\,\mathrm{Re}\,\mathbf{H} \qquad (4.4.2)$$

which can be constructed in a natural way from the electromagnetic tensor $F_{\mu\nu}$ (as is done in the cited texts). The real and imaginary parts of $\varphi_{\mathbf{V}} \equiv \mathbf{V} \cdot \mathbf{V}$ are Lorentz invariant, so the C lines have special invariant status in the field. The (instantaneous) Poynting vector (electromagnetic current density vector) $\mathbf{S}_{\mathrm{Poy}}$ is defined [BW59]

$$\mathbf{S}_{\mathrm{Poy}} = \mathrm{Re}\,\mathbf{E} \wedge \mathrm{Re}\,\mathbf{H}, \qquad (4.4.3)$$

and is the analogue of $\mathbf{N}$ (4.3.4) for the electromagnetic field $\mathbf{V}$, that is, L lines are instantaneous current stagnation lines. The Poynting vector is not Lorentz invariant, and it can be shown that there is a Lorentz transformation that can transform the field such that any point in the field lies on such a current stagnation line, provided that point is not on a C line [PR84b].

The main problem with the interpretation of the field $\mathbf{V}$ is that the singularity lines move (at optical frequencies) even in monochromatic waves (although for these the pattern

is time periodic). In optics, it is more usual to consider the time-averaged Poynting vector $\langle \mathbf{S}_{\mathrm{Poy}} \rangle_t$, where,

$$\langle \mathbf{S}_{\mathrm{Poy}} \rangle_t = \frac{1}{2}(\mathbf{E}^* \wedge \mathbf{H} + \mathbf{E} \wedge \mathbf{H}^*), \tag{4.4.4}$$

where $\mathbf{E}, \mathbf{H}$ are now complex (see [BW59]p33). This real vector is the sum of two cross products like (4.4.3), and does not generically vanish along lines (it cannot be written as a single cross product of real vectors). Therefore the time averaged Poynting vector does not have any time-dependent stagnation lines, so are not a vector analogue for dislocations, which are stationary in monochromatic fields.

In chapter 6, three-dimensional vector fields are realised as spin 1 fields, and the polarization is parameterized by the Majorana sphere $\mathcal{M}_2$ (rather than the Poincaré sphere, which is only appropriate for transverse paraxial fields). This suggests the theory of C and L lines in three dimensions may be recast in a formalism evoking the quantum description of light, and for the remainder of this section we shall reformulate the material of section 4.3 in quantum-mechanical terms.

To this end, we observe that for any two (real) three-vectors $\mathbf{c}, \mathbf{d}$, there is the identity

$$\mathbf{c} \wedge \mathbf{d} = -\mathrm{i}(\mathbf{c} \cdot \hat{\mathbf{S}})\mathbf{d}, \tag{4.4.5}$$

where the hermitian vector operator $\hat{\mathbf{S}}$ is a three-dimensional representation of the spin operator with cartesian basis [VMK88] (not to be confused with the Stokes parameters S_0, S_1, S_2, S_3). Its components are vectors operating on $\mathbf{d}$, and are

$$\hat{\mathbf{S}} = (\hat{S}_x, \hat{S}_y, \hat{S}_z) = \left(\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -\mathrm{i} \\ 0 & \mathrm{i} & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & \mathrm{i} \\ 0 & 0 & 0 \\ -\mathrm{i} & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & -\mathrm{i} & 0 \\ \mathrm{i} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \right), \tag{4.4.6}$$

or more succinctly, in terms of the antisymmetric symbol, $S_i = -\mathrm{i}\varepsilon_{ijk}$. (Note the small correction from [BD01c] equation (3.2)). In quantum mechanics, in units of $\hbar$, $\hat{\mathbf{S}}$ satisfies the commutation rules for spin 1 particles,

$$\hat{\mathbf{S}} \wedge \hat{\mathbf{S}} = \mathrm{i}\hat{\mathbf{S}}. \tag{4.4.7}$$

None of the polarization structure is lost if the complex vector $\mathbf{E}$ is considered as a state (parameterised by $\mathbf{r}, t$), represented by the complex unit vector $\mathbf{e}$

$$\mathbf{e} = \frac{\mathbf{E}}{|\mathbf{E}|} \tag{4.4.8}$$

The polarization ellipse is not affected if the overall phase of $\mathbf{e}$ is ignored, that is, if $\mathbf{E}$ is taken to be a state in projective Hilbert space. In Dirac notation, we have

$$\begin{pmatrix} e_x \\ e_y \\ e_z \end{pmatrix} \equiv |\mathbf{e}\rangle, \qquad \begin{pmatrix} e_x^* & e_y^* & e_z^* \end{pmatrix} \equiv \langle\mathbf{e}|, \qquad \langle\mathbf{e}|\mathbf{e}'\rangle = \mathbf{e}^* \cdot \mathbf{e}'. \tag{4.4.9}$$

The local expectation value $\mathbf{S}$ of the spin operator $\hat{\mathbf{S}}$ (taking advantage of the antisymmetry of its components), is

$$\mathbf{S} = \langle\mathbf{e}|\hat{\mathbf{S}}|\mathbf{e}\rangle = \frac{2\mathbf{N}}{|\mathbf{E}|^2}, \tag{4.4.10}$$

The local spin state can therefore be regarded as a vector perpendicular to the polarization ellipse (with a length equal to the Stokes parameter s_3 defined in the plane normal to $\mathbf{S}$). It is interpreted as the *local angular momentum* of $\mathbf{e}$ at $\mathbf{r}$. C and L lines are loci of particular spin values. By the C conditions,

$$S^2 = \mathbf{S} \cdot \mathbf{S} = (-i\mathbf{e}^* \wedge \mathbf{e}) \cdot (-i\mathbf{e}^* \wedge \mathbf{e}) = 1 - |\mathbf{e} \cdot \mathbf{e}| = 1 \quad \text{on a C line.} \tag{4.4.11}$$

As one would expect, C lines therefore correspond to places where the (modulus of) spin expectation is 1, and in fact $|\mathbf{e}\rangle$ is an eigenstate of the operator $\mathbf{n} \cdot \hat{\mathbf{S}}$,

$$\begin{aligned} (\mathbf{n} \cdot \hat{\mathbf{S}})|\mathbf{e}\rangle &= i\mathbf{n} \wedge \mathbf{e} \\ &= \frac{i((\mathbf{p} \wedge \mathbf{q}) \wedge (\mathbf{p} + i\mathbf{q}))}{|\mathbf{p} \wedge \mathbf{q}|(p^2 + q^2)} \\ &= \frac{i([p^2\mathbf{q} - (\mathbf{p} \cdot \mathbf{q})\mathbf{p}] + i[(\mathbf{p} \cdot \mathbf{q})\mathbf{p} - q^2\mathbf{p}])}{|\mathbf{p} \wedge \mathbf{q}|(p^2 + q^2)} \\ &\qquad\qquad \text{(on C line, } |\mathbf{q}| = |\mathbf{p}|, \mathbf{p} \cdot \mathbf{q} = 0) \\ &= \frac{\mathbf{p} + i\mathbf{q}}{2p^2} \\ &= |\mathbf{e}\rangle. \end{aligned} \tag{4.4.12}$$

The sign of the eigenvalue corresponds to the fact that $\mathbf{n}$ points in the direction of the circulation of $\mathbf{e}$.

For L lines, (4.3.8) and (4.4.10) show that the expectation $\mathbf{S} = \mathbf{0}$ on an L line, and indeed $|\mathbf{e}\rangle$ is an eigenstate of the operator $\mathbf{e}^* \cdot \hat{\mathbf{S}}$ with eigenvalue 0. These observations support the photon interpretation of polarization of light fields.

The vector direction of an L line $\mathbf{D}_{\mathrm{L}}$ may therefore be rewritten more transparently,

$$\mathbf{D}_{\mathrm{L}} = \frac{1}{2}N^2 \operatorname{Im}\langle\nabla\mathbf{n}| \wedge (\mathbf{e_f} \cdot \hat{\mathbf{S}})|\nabla\mathbf{n}\rangle, \tag{4.4.13}$$

where the cross product connects the gradients. This type of notation is frequently used in the theory of geometric phases, where the two spaces (configuration space and state space) may be quite different [SW89].

It is possible to extend the quantum-style description to other properties of the field. For instance, the natural definition of the momentum $\mathbf{k}$ (again in units of $\hbar$) of the field $\mathbf{E}$ is the local expectation of the momentum operator on the state $|\mathbf{e}\rangle$, namely

$$\mathbf{k} \equiv -\mathrm{i}\langle\mathbf{e}|\nabla|\mathbf{e}\rangle \qquad (4.4.14)$$

Nye [Nye91] observed that this geometric phase 1-form (Pancharatnam phase difference) connecting the field at neighbouring points, is the natural definition of a propagation direction at a point in the field. It was found by [NH87] by other means. Unlike the wavevector for rays in geometrical optics, $\mathbf{k}$ is nonintegrable, and there is a geometric phase $\gamma(\Gamma)$ defined by the integral of $\mathbf{k}$ round a circuit Γ in $\mathbf{r}$ (see section 1.4). It may be interpreted as the following: take a vector $\mathbf{e}'$ equal to $\mathbf{e}$ at each point $\mathbf{r}$ along the curve Γ apart from a phase determined by parallel transport from the starting point, where $\mathbf{e}' = \mathbf{e}$ exactly, that is, keeping $\langle\mathbf{e}'|\nabla|\mathbf{e}\rangle = 0$. $\gamma(\Gamma)$ is the phase difference between $\mathbf{e}'$ and $\mathbf{e}$ at the end of the circuit, and is independent of the initial phase of $\mathbf{e}$, so works for $\mathbf{e}$ being a vector in projective Hilbert space [AA87]. $\gamma(\Gamma)$ is also the flux through Γ of a 2-form, which is (using the suffix notation of (4.3.12)),

$$\mathbf{B} = \nabla \wedge \mathbf{k} = \mathrm{Im}\langle\nabla\mathbf{e}| \wedge |\nabla\mathbf{e}\rangle = \mathrm{Im}\,\nabla_a \wedge \nabla_b \mathbf{e}_a^* \cdot \mathbf{e}_b. \qquad (4.4.15)$$

There does not seem to be any simple interpretation of $\mathbf{B}$ in terms of the polarization geometry of the field $\mathbf{E}$, although details of the geometric phase derivation for the Majorana sphere $\mathcal{M}_2$ may be found in [Han98d]. Singularities of $\mathbf{B}$, the codimension 3 monopoles associated with geometric phases [Ber84], occur when the field intensity $\mathbf{E}^* \cdot \mathbf{E}$ vanishes (vector field phase singularities where a C line and L line may intersect), with codimension 6.

Places where $\mathbf{k}$ and $\mathbf{n}$ are (anti)parallel (that is, $\mathbf{k} \wedge \mathbf{n} = 0$, and $\mathbf{k}$ is orthogonal to both $\mathbf{p}, \mathbf{q}$) are helicity states, where the momentum direction corresponds to the normal of the ellipse. This situation occurs generically along lines, which provides further codimension 2 structure to $\mathbf{E}$. [NH87] defined a handedness to the ellipse in the field at every point by $\mathrm{sign}\mathbf{k} \cdot \mathbf{n}$, and regions of right and left handedness are separated by the so-called T ('transverse') surface, on which L lines lie and C lines cross (but helicity lines are restricted

to regions of appropriate handedness). We shall not investigate these structures further here.

4.5 Singularity densities in random paraxial vector waves

4.5.1 Random paraxial vector waves

We intend to generalise the paraxial gaussian random waves of section 3.1, equation (3.1.2), and calculate the densities of C points, disclinations and L lines in two dimensions.

Physically, we are adding together many transverse plane waves, with propagation directions all very close to the z-direction, and any longitudinal z component is negligibly small. The spectral distribution of amplitudes is a paraxial one, such as the disk spectrum (3.1.30) or gaussian spectrum (3.1.33). The ring spectrum, is not appropriate here, since the waves are not propagating in the plane. We shall assume that the spectrum is derived paraxially from a monochromatic three-dimensional wave, and factor out time and z dependence.

The isotropic paraxial gaussian random vector wave superposition analogous to (3.1.2) is

$$\mathbf{E}(\mathbf{R}) = \sum_{\mathbf{K}} a_{\mathbf{K}} \mathbf{d}_{\mathbf{K}} \exp(\mathrm{i}\mathbf{K} \cdot \mathbf{R}), \tag{4.5.1}$$

where, as before, $\mathbf{K} = (K_x, K_y)$, and $a_{\mathbf{K}}$ is a complex amplitude with argument ϕ_K uniformly random and modulus ε_K, independent of the direction of $\mathbf{K}$, related to the plane radial spectrum $\Pi(K)$ by (3.1.10). As before, the nth moment of K with respect to $\Pi(K)$ is denoted K_n, and $\Pi(K)$ is normalised such that $K_0 = 1$.

$\mathbf{d}_{\mathbf{K}}$ is a normalised complex polarization vector describing the polarization state of the plane wave component labelled by $\mathbf{K}$. If the components of $\mathbf{d}_{\mathbf{K}}$ are chosen uniformly at random, one finds that the distribution of polarizations is uniform on the Poincaré sphere (and so, by the discussion in section A.4, uniformly random in the circular components of $\mathbf{d}_{\mathbf{K}}$). Using the Poincaré sphere representation, $\mathbf{d}_{\mathbf{K}}$ has cartesian components

$$\mathbf{d}_{\mathbf{K}} = (\cos\alpha_{\mathbf{K}}/2 \exp(-\mathrm{i}\beta_{\mathbf{K}}/2), \sin\alpha_{\mathbf{K}}/2 \exp(\mathrm{i}\beta_{\mathbf{K}}/2)), \tag{4.5.2}$$

where $\alpha_{\mathbf{K}}, \beta_{\mathbf{K}}$ are the polar angles on the Poincaré sphere for the plane wave labelled by $\mathbf{K}$. The ensemble averaging (cf (3.1.4)) involves averaging over all of the random phases

$\phi_{\mathbf{K}}$, and the Poincaré sphere angles $\alpha_{\mathbf{K}}, \beta_{\mathbf{K}}$:

$$\langle \bullet \rangle = \prod_{\mathbf{K}} \underbrace{\frac{1}{2\pi} \int_0^{2\pi} \mathrm{d}\phi_{\mathbf{K}}}_{\text{phase average}} \underbrace{\frac{1}{4\pi} \int_0^{\pi} \mathrm{d}\alpha_{\mathbf{K}} \sin \alpha_{\mathbf{K}} \int_0^{2\pi} \mathrm{d}\beta_{\mathbf{K}}}_{\text{polarization average}} \bullet . \qquad (4.5.3)$$

This implies that the average length of $\mathbf{E}$ is normalised,

$$\langle \mathbf{E}^* \cdot \mathbf{E} \rangle = 1. \qquad (4.5.4)$$

(4.5.1) is a complex vector, and the real and imaginary parts of its components P_x, P_y, Q_x, Q_y, are gaussian random functions of position $\mathbf{R}$, and the ensemble is stationary, ergodic, etc. Moreover, we can write

$$\begin{aligned} \mathbf{P} &= \sum_{\mathbf{K}} \varepsilon_K \mathbf{X_K}, \\ \mathbf{Q} &= \sum_{\mathbf{K}} \varepsilon_K \mathbf{Y_K}, \end{aligned} \qquad (4.5.5)$$

where $\mathbf{X_K}, \mathbf{Y_K}$ are found from (4.5.2) (suppressing $\mathbf{K}$ suffices),

$$\begin{aligned} \mathbf{X} &= (\cos \alpha/2 \cos(\mathbf{K} \cdot \mathbf{R} + \phi - \beta/2), \sin \alpha/2 \cos(\mathbf{K} \cdot \mathbf{R} + \phi + \beta/2)), \\ \mathbf{Y} &= (\cos \alpha/2 \sin(\mathbf{K} \cdot \mathbf{R} + \phi - \beta/2), \sin \alpha/2 \sin(\mathbf{K} \cdot \mathbf{R} + \phi + \beta/2)). \end{aligned} \qquad (4.5.6)$$

The components of these vectors are averaged to give

$$\langle X_i^2 \rangle = \langle Y_i^2 \rangle = \frac{1}{4}, \qquad \langle X_i X_j \rangle = \langle Y_i Y_j \rangle = \langle X_i Y_j \rangle = \langle X_i Y_i \rangle = 0, \qquad i \neq j \text{ are } x, y, \qquad (4.5.7)$$

implying that the gaussian random functions P_x, P_y, Q_x, Q_y are all identically and independently distributed. Using the notation $\rho(f)$ for the probability density function of random variable f (since the symbol P, used in chapter 3, is in use), the probability density function $\rho(\mathbf{E})$ is

$$\rho(\mathbf{E}) = \frac{4}{\pi^2} \exp(-2\mathbf{E}^* \cdot \mathbf{E}) = \frac{4}{\pi^2} \exp(-2(P_x^2 + P_y^2 + Q_x^2 + Q_y^2)). \qquad (4.5.8)$$

Also, from (4.5.6),

$$\partial_i X_j = -K_i Y_j, \qquad \partial_i Y_j = K_i X_j, \qquad i, j = x, y, \qquad (4.5.9)$$

so the derivative fields $P_{i,j}, Q_{i,j}$ are also gaussian random, whose only nonvanishing correlations are

$$\langle P_{i,j}^2 \rangle = \langle Q_{i,j}^2 \rangle = \frac{K_2}{8}, \qquad i, j = x, y. \qquad (4.5.10)$$

These functions can be used to determine the probability density functions of the unnormalised Stokes parameters S_0, S_1, S_2, S_3, which are

$$\rho(S_0) = 4S_0 \exp(-2S_0), \qquad \rho(S_i) = \exp(-2|S_i|), \qquad i = 1, 2, 3, \qquad (4.5.11)$$

(the normalised Stokes parameters s_1, s_2, s_3 are uniformly distributed between 0 and 1), and these calculations agree with those of [Bar87, Bro95] in the case of completely unpolarized radiation.

4.5.2 Density of paraxial C points

Since the random wavefield (4.5.1) is statistically stationary and ergodic, and C points are phase singularities of the polarization scalar $\varphi = u + \mathrm{i}v$ (4.2.6), the density of C points $d_{C,2}$ can be found by direct analogy with (3.2.1):

$$d_{C,2} = \langle \delta(u)\delta(v)|u_x v_y - u_y v_x| \rangle. \qquad (4.5.12)$$

Writing the jacobian $|u_x v_y - u_y v_x| \equiv |J|$, the average is

$$d_{C,2} = \int \mathrm{d}^2\mathbf{P}\mathrm{d}^2\mathbf{Q}\mathrm{d}^4\nabla\mathbf{P}\mathrm{d}^4\nabla\mathbf{Q} \, \delta(P^2 - Q^2)\delta(2\mathbf{P}\cdot\mathbf{Q}) \, |J| \, \rho(\mathbf{P},\mathbf{Q},\nabla\mathbf{P},\nabla\mathbf{Q}). \qquad (4.5.13)$$

We shall integrate $\mathbf{P}, \mathbf{Q}$ first, and write the integrals involving the first derivatives as I_1. Rewriting $\mathbf{P}, \mathbf{Q}$ in polar coordinates $(P, \theta_0), (Q, \theta_0 + \theta)$ and integrating θ_0 out easily,

$$d_{C,2} = \frac{8}{\pi} \int_0^\infty \mathrm{d}P \int_0^\infty \mathrm{d}Q \int_0^{2\pi} \mathrm{d}\theta \, PQ \, \delta(P^2 - Q^2)\delta(2PQ\cos\theta) \exp(-2(P^2 + Q^2)) \, I_1. \qquad (4.5.14)$$

The δ-functions are easy to integrate (with respect to Q and θ), and the integral becomes

$$d_{C,2} = \frac{16}{\pi} \int_0^\infty \mathrm{d}P \frac{P^2}{4P^3} \exp(-4P^2) I_{1(Q=P,\cos\theta=0)} \qquad (4.5.15)$$

and the notation $I_{1(Q=P,\cos\theta=0)}$ is obvious. The jacobian $|J|$ now simplifies:

$$|J| = 4P^2|P_{x,x}P_{y,y} + P_{x,x}Q_{x,y} - P_{y,y}Q_{y,x} + Q_{x,x}Q_{y,y} - P_{x,y}P_{y,x} - P_{x,y}Q_{x,x}$$
$$- Q_{x,y}Q_{y,x} + P_{y,x}Q_{y,y}|. \quad (4.5.16)$$

Rewriting the first derivatives as a vector

$$\mathbf{V} = \sqrt{K_2}/2(P_{x,x}, P_{x,y}, P_{y,x}, P_{y,y}, Q_{x,x}, Q_{x,y}, Q_{y,x}, Q_{y,y})$$

the jacobian in (4.5.16) can be written as a quadratic form, and after rescaling (to remove factors in the gaussian), is equal to $(K_2/4)^4 |\mathbf{V} \cdot \mathbf{\Xi} \cdot \mathbf{V}|$, with $\mathbf{\Xi}$ numerical and symmetric. Therefore

$$d_{\mathrm{C},2} = \frac{16}{\pi} \left(\frac{4}{\pi K_2} \right)^4 \int_0^\infty \mathrm{d}P \, P \exp(-4P^2) \int \mathrm{d}^8 \mathbf{V} \, |\mathbf{V} \cdot \mathbf{\Xi} \cdot \mathbf{V}| \exp(-V^2). \qquad (4.5.17)$$

Now we perform a linear transformation $\mathbf{V} \to \mathbf{W} = \mathbf{\Gamma V}$ with $\mathbf{\Gamma}$ an orthogonal matrix diagonalising $\mathbf{\Xi}$,

$$\mathbf{\Gamma}^T \mathbf{\Xi} \mathbf{\Gamma} = \mathrm{diag}\{1, 1, -1, -1, 0, 0, 0, 0\}. \qquad (4.5.18)$$

Such a $\mathbf{\Gamma}$ can always be found because $\mathbf{\Xi}$ is real symmetric. The jacobian of this transformation is 1 since $\mathbf{\Gamma}$ is orthogonal, so after the transformation, the calculation concludes:

$$
\begin{aligned}
d_{\mathrm{C},2} &= \frac{K_2}{2\pi^5} \int \mathrm{d}^8 \mathbf{W} \, |W_1^2 + W_2^2 - W_3^2 - W_4^2| \exp(-W^2) \\
&\qquad \text{(transforming to plane polars, } (W_1, W_2) \to (r_1, \phi_1), (W_3, W_4) \to (r_2, \phi_2), \\
&\qquad\qquad \text{then integrating out angles and } W_5, \dots W_8) \\
&= \frac{K_2}{2\pi^5} 4\pi^4 \int_0^\infty \mathrm{d}r_1 \int_0^\infty \mathrm{d}r_2 \, r_1 r_2 |r_1^2 - r_2^2| \exp(-r_1^2 - r_2^2) \\
&\qquad\qquad \text{(transforming to polars } (r_1, r_2) \to (r, \phi)) \\
&= \frac{2K_2}{\pi} \int_0^\infty \mathrm{d}r \, r^5 \exp(-r^2) \int_0^{\pi/2} \mathrm{d}\phi \, \cos\phi \sin\phi |\cos 2\phi| \\
&= \frac{K_2}{2\pi}. \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (4.5.19)
\end{aligned}
$$

The density of paraxial C points is therefore twice the planar dislocation density (3.2.2). This is not surprising, as each type of C point (right, left handed) is a dislocation in a circular component α_+, α_-; each of these components acts just like a random scalar field.

4.5.3   Density of paraxial L lines

By stationarity and ergodicity, the average length of L line per unit volume can be found from (4.3.15), giving

$$d_{\mathrm{L},2} = \langle \delta(N) |\nabla N| \rangle. \qquad (4.5.20)$$

where $N = P_x Q_y - P_y Q_x$ from (4.2.4). Now, $\mathbf{P}$ and $\mathbf{Q}$ may be transformed to polar coordinates as before (but now, by isotropy, we choose $\mathbf{P}$ to be $(P, 0)$), and

$$
\begin{aligned}
d_{\mathrm{L},2} &= \frac{8}{\pi} \int_0^\infty \mathrm{d}P \int_0^\infty \mathrm{d}Q \int_0^{2\pi} \mathrm{d}\theta \, PQ\delta(PQ\sin\theta)\exp(-2(P^2+Q^2)) \\
&\quad \times \int \mathrm{d}\nabla\mathbf{P}\mathrm{d}\nabla\mathbf{Q}\,\rho(\nabla\mathbf{P},\nabla\mathbf{Q})|\nabla N| \\
&= \frac{16}{\pi} \int_0^\infty \mathrm{d}P \int_0^\infty \mathrm{d}Q\exp(-2(P^2+Q^2)) \\
&\quad \times \int \mathrm{d}\nabla\mathbf{P}\mathrm{d}\nabla\mathbf{Q}\,\rho(\nabla\mathbf{P},\nabla\mathbf{Q})|\nabla N|.
\end{aligned}
\tag{4.5.21}
$$

Transforming to polars $(P, Q) \to (T, \phi)$, and using the fact that both $\mathbf{P}, \mathbf{Q}$ are in the x-direction, $|\nabla N|$ becomes

$$
T\sqrt{c^2(\nabla Q_y)^2 + s^2(\nabla P_y)^2 - 2cs\nabla P_y \cdot \nabla Q_y},
\tag{4.5.22}
$$

where c, s denote $\cos\phi, \sin\phi$ respectively. $|\nabla N|$ only now involves the derivatives of the y components of $\mathbf{P}, \mathbf{Q}$, so only these need to be integrated over. Writing these as a vector $\mathbf{V} = \sqrt{K_2}/2(P_{y,x}, P_{y,y}, Q_{y,x}, Q_{y,y})$, and writing $|\nabla N|$ as a quadratic form $TK_2/4|\mathbf{V}\cdot\mathbf{\Xi}\cdot\mathbf{V}|^{1/2}$, the integral becomes

$$
d_{\mathrm{L},2} = \frac{16}{\pi} \int_0^\infty \mathrm{d}T\, T^2 \exp(-2T^2) \int_0^\infty \mathrm{d}\phi \frac{\sqrt{K_2/2}}{4\pi^2} \int \mathrm{d}^4\mathbf{V}|\mathbf{V}\cdot\mathbf{\Xi}\cdot\mathbf{V}|\exp(-V^2).
\tag{4.5.23}
$$

As before, $\mathbf{V}$ can be orthogonally transformed to a basis in which $\mathbf{\Xi}$ is diagonal, and therefore can easily be integrated, with the result

$$
d_{\mathrm{L},2} = \frac{\pi}{4}\sqrt{\frac{K_2}{2}}.
\tag{4.5.24}
$$

The density of L lines per unit area is related to the density $d_{\mathrm{L},1}$ of (point) intersections of L lines with a straight line in the plane, by the same argument as that used in section 3.3.1. The two densities differ by a factor equal to the average modulus of x-component of a random isotropic unit vector, ie a factor of $2/\pi$, so

$$
d_{\mathrm{L},1} = \frac{1}{2}\sqrt{\frac{K_2}{2}}.
\tag{4.5.25}
$$

4.5.4 Paraxial disclination density

We now compute the disclination density, which is easily done by choosing the phase $\chi = 0$; we are finding the density of zeros of the real vector field $\mathbf{P}$. By ergodicity from

Singularity type	General value	Disk spectrum	Gaussian spectrum
Dislocation density $d_{D,2}$	$\frac{K_2}{4\pi}$	$\frac{\pi}{2\Lambda_{\mathrm{d}}^2}$	$\frac{2\pi}{\Lambda_\sigma^2}$
C point density $d_{\mathrm{C},2}$	$\frac{K_2}{2\pi}$	$\frac{\pi}{\Lambda_{\mathrm{d}}^2}$	$\frac{4\pi}{\Lambda_\sigma^2}$
L line density $d_{\mathrm{L},2}$	$\frac{\pi}{4}\sqrt{\frac{K_2}{2}}$	$\frac{\pi^2}{4\Lambda_{\mathrm{d}}}$	$\frac{\pi^2}{2\Lambda_\sigma}$
Disclination density $\quad d_{\mathrm{disc},\,2}$	$\frac{K_2}{4\pi}$	$\frac{\pi}{2\Lambda_{\mathrm{d}}^2}$	$\frac{2\pi}{\Lambda_\sigma^2}$

Table 4.1: Statistical densities of paraxial polarization singularities and comparison with planar dislocation density, for general values, the disk spectrum and the gaussian spectrum.

(4.2.22), the disclination density $d_{\mathrm{disc},2}$ is

$$
\begin{aligned}
d_{\mathrm{disc},\,2} &= \langle \delta^2(\mathbf{P})|P_{x,x}P_{y,y} - P_{x,y}P_{y,x}|\rangle \\
&= \int \mathrm{d}^2\mathbf{P}\mathrm{d}^4\nabla\mathbf{P}\delta(P_x)\delta(P_y)|P_{x,x}P_{y,y} - P_{x,y}P_{y,x}|\rho(\mathbf{P},\nabla\mathbf{P}) \\
&= \frac{K_2}{4\pi}
\end{aligned}
\tag{4.5.26}
$$

the same as the paraxial dislocation density, and half the C point density. Formally, this calculation is identical to that for two dimensional dislocation density (3.2.2) since the statistical relationship between P_x, P_y is the same as that between ξ, η; it is also given by [Hal81] equation (6.25). The mean spacing of disclinations on L lines is $d_{\mathrm{L},2}/d_{\mathrm{disc},2} = \pi^2/\sqrt{2K_2}$.

The results of this section are summarised in table (4.1), where values of the density are also given for the disk and gaussian spectra.

4.6 Singularity densities in random spatial vector waves

4.6.1 The random three-dimensional wave model

We now consider the three-dimensional analogue to the paraxial vector waves of the last section. As with three-dimensional random scalar waves, the ensemble is made up of superpositions of an infinite number of plane waves with random phases (each randomly polarized transverse to its propagation direction), with propagations uniformly distributed in direction; the normal $\mathbf{n}(\mathbf{r})$ to the resulting polarization ellipse at any point $\mathbf{r}$ changes smoothly over space. Although the random waves constructed shall be the complex analytic signal of a real complex wavefield, we shall see the statistical model applies equally well to the electromagnetic field $\mathbf{V} = \mathrm{Re}\,\mathbf{E} + \mathrm{i}\,\mathrm{Im}\,\mathbf{H}$ of equation (4.4.2).

By analogy with (3.1.1), (4.5.1), the isotropic random complex three-dimensional vector wave superposition $\mathbf{E}(\mathbf{r})$ is (ignoring t dependence)

$$\mathbf{E}(\mathbf{r}) = \sum_{\mathbf{k}} a_{\mathbf{k}} \mathbf{d}_{\mathbf{k}} \exp(i\mathbf{k} \cdot \mathbf{r}), \qquad (4.6.1)$$

where each wave in the superposition is labelled by its wavevector $\mathbf{k}$. As before, the $a_{\mathbf{k}}$ are complex scalar amplitudes with uniformly distributed phases $\phi_{\mathbf{k}}$ and moduli ε_k related to the spatial radial power spectrum $\Pi(k)$ by (3.1.8). Moments of k with respect to $\Pi(k)$ are denoted k_n ($k_0 = 0$), as with (3.1.9). The polarization vector $\mathbf{d}_{\mathbf{k}}$, is defined in the plane orthogonal to $\mathbf{k}$ identically to the planar case. The right-handed orthogonal frame $(\mathbf{u}_{\mathbf{k}}, \mathbf{v}_{\mathbf{k}}, \mathbf{k})$ with $u_{\mathbf{k}}^2 = v_{\mathbf{k}}^2 = 1$, and

$$\mathbf{d}_{\mathbf{k}} = \mathbf{u}_{\mathbf{k}} \cos \alpha_{\mathbf{k}} \exp(-i\beta_{\mathbf{k}}/2) + \mathbf{v}_{\mathbf{k}} \sin \alpha_{\mathbf{k}}/2 \exp(i\beta_{\mathbf{k}}/2) \qquad (4.6.2)$$

where $\alpha_{\mathbf{k}}, \beta_{\mathbf{k}}$ are polar angles on the Poincaré sphere defined in the $(\mathbf{u}_{\mathbf{k}}, \mathbf{v}_{\mathbf{k}})$ plane, chosen uniformly randomly as before. By virtue of the central limit theorem, $\mathbf{E}$ in (4.6.1) represents an ensemble of complex gaussian random vectors, parameterised by the random $\phi_{\mathbf{k}}, \alpha_{\mathbf{k}}, \beta_{\mathbf{k}}$, with ensemble averaging analogous to (4.5.3). As with paraxial waves, these conditions imply that, on the average, $\mathbf{E}$ is normalised, as with (4.5.4).

The frame $(\mathbf{u}_{\mathbf{k}}, \mathbf{v}_{\mathbf{k}}, \mathbf{k})$ is chosen such that, if $\mathbf{w}_{\mathbf{k}} = \mathbf{k}/k$ is the direction vector of $\mathbf{k}$ in space, parameterised by polar angles θ, ϕ, then

$$\begin{aligned} \mathbf{w}_{\mathbf{k}} &= (\cos\phi \sin\theta, \sin\phi \sin\theta, \cos\theta), \\ \mathbf{v}_{\mathbf{k}} &= \frac{\mathbf{e}_z \wedge \mathbf{w}_{\mathbf{k}}}{|\mathbf{e}_z \wedge \mathbf{w}_{\mathbf{k}}|} = (-\sin\phi, \cos\phi, 0), \\ \mathbf{u}_{\mathbf{k}} &= \mathbf{v}_{\mathbf{k}} \wedge \mathbf{w}_{\mathbf{k}} = (\cos\phi \cos\theta, \sin\phi \cos\theta, -\sin\theta). \end{aligned} \qquad (4.6.3)$$

(The weighting when $\mathbf{w}_{\mathbf{k}} = \pm\mathbf{e}_z$, and $\mathbf{v}_{\mathbf{k}}, \mathbf{u}_{\mathbf{k}}$ are singular, is negligible.) The real and imaginary parts of $\mathbf{E}$ are

$$\mathbf{p} = \sum_{\mathbf{k}} \varepsilon_k \mathbf{X}_{\mathbf{k}}, \qquad (4.6.4)$$

$$\mathbf{q} = \sum_{\mathbf{k}} \varepsilon_k \mathbf{Y}_{\mathbf{k}}, \qquad (4.6.5)$$

where $\mathbf{X}_{\mathbf{k}}, \mathbf{Y}_{\mathbf{k}}$ are defined

$$\begin{aligned} \mathbf{X}_{\mathbf{k}} &= \mathbf{u}_{\mathbf{k}} \cos \alpha_{\mathbf{k}}/2 \cos(\mathbf{k} \cdot \mathbf{r} + \phi_{\mathbf{k}} - \beta_{\mathbf{k}}/2) + \mathbf{v}_{\mathbf{k}} \sin \alpha_{\mathbf{k}}/2 \cos(\mathbf{k} \cdot \mathbf{r} + \phi_{\mathbf{k}} + \beta_{\mathbf{k}}/2), \\ \mathbf{Y}_{\mathbf{k}} &= \mathbf{u}_{\mathbf{k}} \cos \alpha_{\mathbf{k}}/2 \sin(\mathbf{k} \cdot \mathbf{r} + \phi_{\mathbf{k}} - \beta_{\mathbf{k}}/2) + \mathbf{v}_{\mathbf{k}} \sin \alpha_{\mathbf{k}}/2 \sin(\mathbf{k} \cdot \mathbf{r} + \phi_{\mathbf{k}} + \beta_{\mathbf{k}}/2) \end{aligned} \qquad (4.6.6)$$

(cf (4.5.6), and note that the phases $\phi_{\mathbf{k}}$ are unrelated to the spatial azimuth ϕ). In the $\mathbf{u_k}, \mathbf{v_k}$ plane, the components of $\mathbf{X_k}, \mathbf{Y_k}$, satisfy (4.5.7). Averaging the components of $\mathbf{p}$ (the result for $\mathbf{q}$ is the same), and suppressing obvious subscripts $\mathbf{k}$ on $\mathbf{u_k}, \mathbf{v_k}$,

$$
\begin{aligned}
\langle p_i p_j \rangle &= \sum_{\mathbf{k}} \varepsilon_k^2 (\langle X_{\mathbf{k}1}^2 \rangle u_i u_j + \langle X_{\mathbf{k}2}^2 \rangle v_i v_j + \langle X_{\mathbf{k}1} X_{\mathbf{k}2} \rangle (u_i v_j + u_j v_i)) \\
&= \frac{1}{4} \frac{1}{4\pi} \underbrace{\int_0^\pi \mathrm{d}\theta \int_0^{2\pi} \mathrm{d}\phi (u_i u_j + v_i v_j)}_{\frac{2}{3} 4\pi \delta_{ij}} \int_0^\infty \mathrm{d}k \, \Pi(k) \\
&= \frac{\delta_{ij}}{6},
\end{aligned}
\tag{4.6.7}
$$

(where $i, j = x, y, z$, and δ_{ij} is the Kronecker δ-symbol) agreeing with the statistical normalisation of $\mathbf{E}$. By similar arguments, it can be shown

$$
\langle p_i q_j \rangle = 0.
\tag{4.6.8}
$$

Therefore $\mathbf{p}, \mathbf{q}$ have the joint probability density function

$$
\rho(\mathbf{p}, \mathbf{q}) = \left(\frac{3}{\pi} \right)^3 \exp(-3(p^2 + q^2)).
\tag{4.6.9}
$$

Using the three dimensional analogue to (4.5.9), the derivatives of $\mathbf{p}, \mathbf{q}$ have the nonvanishing averages

$$
\begin{aligned}
\langle p_{i,i}^2 \rangle = \langle q_{i,i}^2 \rangle &= \frac{k_2}{30} \\
\langle p_{i,j}^2 \rangle = \langle q_{i,j}^2 \rangle &= \frac{k_2}{15} \\
\langle p_{i,j} p_{j,i} \rangle = \langle q_{i,j} q_{j,i} \rangle &= -\frac{k_2}{60}.
\end{aligned}
\tag{4.6.10}
$$

Since $\mathbf{E}$ is divergenceless (being a sum of plane waves), the random variables $p_{x,x}, p_{y,y}, p_{z,z}$ are dependent, and in fact, in all calculations, $p_{z,z}$ is replaced by $-p_{x,x} - p_{y,y}$, with equivalent substitution for $\mathbf{q}$.

The distributions of $\mathbf{p}, \mathbf{q}$ can be used to find the probability density function of the square of local angular momentum expectation value $\mathbf{S}$, defined in equation (4.4.10).

Therefore

$$
\begin{aligned}
\rho(S^2) &= \left\langle \delta\left(S^2 - \left(\frac{2\mathbf{p} \wedge \mathbf{q}}{p^2 + q^2}\right)^2\right)\right\rangle \\
&= \left(\frac{3}{\pi}\right)^3 \int d^3\mathbf{p} \int d^3\mathbf{q} \exp(-3(p^2 + q^2))\delta\left(S^2 - \left(\frac{2\mathbf{p} \wedge \mathbf{q}}{p^2 + q^2}\right)^2\right) \\
&\qquad \text{(exploiting isotropy, transforming } \mathbf{p}, \mathbf{q} \text{ to spherical polars)} \\
&= \frac{216}{\pi} \int_0^\infty dp \int_0^\infty dq \int_0^\pi d\theta\, p^2 q^2 \sin\theta \exp(-3(p^2 + q^2))\delta\left(S^2 - \frac{4p^2 q^2 \sin^2\theta}{(p^2 + q^2)^2}\right) \\
&\qquad \text{(transforming } (p, q) \text{ to plane polars } (r, \phi)) \\
&= \frac{108}{\pi} \int_0^\infty r^5 \exp(-3r^2) \int_0^{\pi/2} d\phi \int_0^{\pi/2} d\theta \sin\theta \sin^2(2\phi)\delta(S^2 - \sin^2 2\phi \sin^2\theta) \\
&= 1. \tag{4.6.11}
\end{aligned}
$$

The distribution of S^2 is therefore uniform in space, between values of 0 and 1.

For the remainder of this section, we shall be concerned with the electromagnetic vector $\mathbf{V}$ of (4.4.2), with $\mathbf{E}, \mathbf{H}, \mathbf{A}$ real. Let $\mathbf{E}$ be equal to $\mathbf{p}$, defined in equation (4.6.4), with time dependence included in the obvious way, by replacing $\mathbf{k} \cdot \mathbf{r}$ by $\mathbf{k} \cdot \mathbf{r} - \omega_k t$. The real vector potential $\mathbf{A}$ is such that $\partial_t \mathbf{A} = -\mathbf{E}$, so integrating,

$$
\mathbf{A}(\mathbf{r}, t) = -\sum_{\mathbf{k}} \frac{\varepsilon_k}{\omega_k} \mathbf{Y_k} \tag{4.6.12}
$$

and $\nabla \wedge \mathbf{A} = \mathbf{H}$ gives

$$
\mathbf{H}(\mathbf{r}, t) = -\frac{1}{\mu_0 c} \sum_{\mathbf{k}} \frac{\varepsilon_k}{\omega_k} \mathbf{Z_k} \tag{4.6.13}
$$

where

$$
\mathbf{Z_k} = \mathbf{v_k} \cos\alpha_{\mathbf{k}}/2 \cos(\mathbf{k} \cdot \mathbf{r} + \phi_{\mathbf{k}} - \beta_{\mathbf{k}}/2) - \mathbf{u_k} \sin\alpha_{\mathbf{k}}/2 \cos(\mathbf{k} \cdot \mathbf{r} + \phi_{\mathbf{k}} + \beta_{\mathbf{k}}/2). \tag{4.6.14}
$$

Comparison of $\mathbf{X_k}$ and $\mathbf{Z_k}$ confirms that, for each individual plane wave component $\mathbf{k}$, $\mathbf{E_k} \cdot \mathbf{H_k} = 0$ and $\mathbf{E_k} \wedge \mathbf{H_k} \,||\, \mathbf{k}$.

All components of $\mathbf{E}$ are independent of those of $\mathbf{H}$, since

$$
\langle E_i H_j \rangle = \frac{1}{4\mu_0 c} \sum \varepsilon_k^2 (u_{\mathbf{k}i} v_{\mathbf{k}j} - u_{\mathbf{k}i} v_{\mathbf{k}j}), \tag{4.6.15}
$$

which disappears for each choice of $i, j = x, y, z$ on integration in $\mathbf{k}$-space. The other statistics are similar to those derived above.

4.6.2   Density of C lines

Only an outline of the calculation of the C line density $d_{C,3}$ is given here (full details are given in [BD01c]). The result is discussed in section 4.6.4.

Where u, v are now the real and imaginary parts of the polarization scalar $\varphi = \mathbf{E} \cdot \mathbf{E}$ (4.2.6), the density of C lines in space $d_{C,3}$ is given by an expression analogous to (3.3.1),

$$d_{C,3} = \langle \delta(u)\delta(v)|\nabla u \wedge \nabla v| \rangle. \tag{4.6.16}$$

For now, let $\mathbf{U}, \mathbf{V}$, be $\nabla u, \nabla v$ respectively. We begin by finding the conditional probability density function $\rho(\mathbf{U}, \mathbf{V}; \mathbf{p}, \mathbf{q})$ of $\mathbf{U}, \mathbf{V}$ with respect to certain fixed (but arbitrary) values of $\mathbf{p}$ and $\mathbf{q}$. In an obvious notation

$$
\begin{aligned}
\rho(\mathbf{U}, \mathbf{V}; \mathbf{p}, \mathbf{q}) &= \langle \delta(\mathbf{U} - \nabla u)\delta(\mathbf{V} - \nabla v) \rangle_{(\mathbf{p},\mathbf{q})} \\
&= \left(\frac{1}{2\pi}\right)^6 \int d^3s d^3t \, \exp(i(\mathbf{U} \cdot \mathbf{s} + \mathbf{V} \cdot \mathbf{t})) \langle \exp(-i(\mathbf{s} \cdot \nabla u + \mathbf{t} \cdot \nabla v)) \rangle_{(\mathbf{p},\mathbf{q})} \\
&= \left(\frac{1}{2\pi}\right)^6 \int d^3s d^3t \, \exp(i(\mathbf{U} \cdot \mathbf{s} + \mathbf{V} \cdot \mathbf{t})) \exp(-T/2), \tag{4.6.17}
\end{aligned}
$$

by (3.1.22), (3.1.24); summing repeated indices,

$$T \equiv 4[(s_i s_k + t_i t_k)(p_j p_l + q_j q_l)\langle p_{j,i} p_{k,l} \rangle]. \tag{4.6.18}$$

Thus

$$
d_{C,3} = \left(\frac{3}{\pi}\right)^3 \int d^3\mathbf{p} \int d^3\mathbf{q} \, \delta(p^2 - q^2)\delta(2\mathbf{p} \cdot \mathbf{q}) \exp(-3(p^2 + q^2))
$$
$$
\times \int d^3\mathbf{U} \int d^3\mathbf{V} |\mathbf{U} \wedge \mathbf{V}| \rho(\mathbf{U}, \mathbf{V}; \mathbf{p}, \mathbf{q}). \tag{4.6.19}
$$

Isotropy is now used to simplify the $\mathbf{U}, \mathbf{V}$ integrals, by choosing $\mathbf{p}_f = (p, 0, 0)$. Integrating the δ-functions, we put $\mathbf{q}_f = (0, p, 0)$, and find that (4.6.18) becomes

$$T = 4p^2 k_2 \left[\frac{1}{10}(s_x^2 + s_y^2 + t_x^2 + t_y^2) + \frac{2}{15}(s_z^2 + t_z^2)\right]. \tag{4.6.20}$$

The $\mathbf{s}, \mathbf{t}$ integrations in (4.6.17) are easy gaussians, and after rescaling $\mathbf{U}, \mathbf{V}$ to remove factors of p and k_2 from the exponential, the $\mathbf{p}, \mathbf{q}$ integrals in (4.6.19) are easy and similar to those in (4.5.14), (4.5.15), leaving the integral

$$d_{C,3} = \frac{9k_2}{20\pi^4} \int d^3\mathbf{U} \int d^3\mathbf{V} |\mathbf{U} \wedge \mathbf{V}| \exp(-[U^2 + V^2 - \frac{1}{4}(U_z^2 + V_z^2)]) \tag{4.6.21}$$

Equation (3.2.9), may be generalised to find a Fourier expression for the modulus of a 3-vector $\mathbf{W}$,

$$|\mathbf{W}| = -\frac{1}{2\pi} \int \frac{\mathrm{d}^3 \mathbf{t}}{t^2} \nabla_{\mathbf{t}}^2 \exp(i\mathbf{W} \cdot \mathbf{t}). \tag{4.6.22}$$

This may be applied to the vector product term in (4.6.21) to give a 6×6 quadratic form matrix $\mathbf{M}$ (given explicitly in [BD01c] eq (C4)), with determinant

$$\det \mathbf{M} = \frac{1}{16}(3 + 4(t_1^2 + t_2^2) + 3t_3^2)^2. \tag{4.6.23}$$

The $\mathbf{U}, \mathbf{V}$ integrals are an easy 6-dimensional gaussian vector integral, giving

$$d_{\mathrm{C},3} = -\frac{9k_2}{5\pi^3} \int \frac{\mathrm{d}^3 \mathbf{t}}{t^2} \nabla_{\mathbf{t}}^2 \frac{1}{\sqrt{\det \mathbf{M}}}. \tag{4.6.24}$$

The laplacian is easily evaluated, and in the resulting integral the vector $\mathbf{t}$ is naturally expressed in terms of cylindrical coordinates R_t, ϕ_t, and $z_t \equiv t$. The azimuthal and radial integrals follow, and the final integral is written

$$d_{\mathrm{C},3} = -\frac{72k_2}{5\pi^2} \int \mathrm{d}t[g(t) + h(t)], \tag{4.6.25}$$

where

$$g(t) = \frac{19 + 16t^2 + 5t^4}{(t^2 - 3)^3(t^2 + 1)^2}, \qquad h(t) = \frac{33 + 17t^2}{(t^2 - 3)^2} \log\left(\frac{3(1 + t^2)}{4t^2}\right). \tag{4.6.26}$$

These are tricky but standard integrals that can be integrated by standard methods of complex contour integration. The final answer is

$$d_{\mathrm{C},3} = k_2\left(\frac{3}{10\pi} + \frac{1}{5\sqrt{3}}\right) = 0.21096k_2. \tag{4.6.27}$$

This result is discussed below in 4.6.4.

4.6.3 Density of L lines

This section, like the last one, is an outline of the derivation of the L line density $d_{\mathrm{L},3}$ in space, where details of the calculation can be found in [BD01c], and discussion in section 4.6.4.

As before, taking advantage of ergodicity, the density of L lines is, in analogy with (4.3.15),

$$d_{\mathrm{L},3} = \left\langle \frac{\delta(|\mathbf{p} \wedge \mathbf{q}|)}{\pi|\mathbf{p} \wedge \mathbf{q}|}|\mathbf{D}_{\mathrm{L}}| \right\rangle \tag{4.6.28}$$

where $\mathbf{D}_{\mathrm{L}}$ is given in (4.3.12). The calculation is performed in a similar way to that of $d_{\mathrm{C},3}$; $\mathbf{p}$ is fixed to $(0,0,p)$, with jacobian 4π, and on the L line, $\mathbf{D}_{\mathrm{L}}$ becomes $\mathbf{A} \wedge \mathbf{B}$ as in (4.3.11). By analogy with (4.6.17), we write the conditional probability density function (anticipating the L-condition in the δ-function giving $\mathbf{q} = (0,0,q)$),

$$
\begin{aligned}
\rho(\mathbf{A}, \mathbf{B}; \mathbf{p}, \mathbf{q}) &= \langle \delta(\mathbf{A} + p\nabla q_y - q\nabla p_y)\delta(\mathbf{B} - p\nabla q_x - q\nabla p_y)\rangle_{(\mathbf{p},\mathbf{q})} \\
&= \frac{1}{(2\pi)^6} \int \mathrm{d}^3\mathbf{s}\,\mathrm{d}^3\mathbf{t}\, \exp(\mathrm{i}(\mathbf{A} \cdot \mathbf{s} + \mathbf{B} \cdot \mathbf{t}) - F/2) \quad (4.6.29)
\end{aligned}
$$

where F is given by

$$
F = \langle [\mathbf{s} \cdot (-p\nabla q_y + q\nabla p_y) + \mathbf{t} \cdot (p\nabla q_x - q\nabla p_x)]^2 \rangle_{(\mathbf{p},\mathbf{q})}. \quad (4.6.30)
$$

(4.6.28) can now be rewritten, with $\mathbf{p}$ already fixed, and the azimuthal direction of $\mathbf{q}$ fixed with jacobian 2π, with $\mathbf{p} \cdot \mathbf{q} = \cos\theta$,

$$
\begin{aligned}
d_{\mathrm{L},3} &= \left(\frac{3}{\pi}\right)^3 8\pi^2 \int_0^\infty \mathrm{d}p \int_0^\infty \mathrm{d}q \int_0^\pi \mathrm{d}\theta\, p^2 q^2 \sin\theta \frac{\delta(pq\sin\theta)}{\pi pq\sin\theta} \exp(-3(p^2+q^2)) \\
&\quad \times \int \mathrm{d}^3\mathbf{A}\,\mathrm{d}^3\mathbf{B}\, |\mathbf{A} \wedge \mathbf{B}| \rho(\mathbf{A}, \mathbf{B}; \mathbf{p}, \mathbf{q}) \\
&= \frac{216}{\pi^2} \int_0^\infty \mathrm{d}p \int_0^\infty \mathrm{d}q\, \exp(-3(p^2+q^2)) \int_0^\pi \mathrm{d}\theta\, \delta(\sin\theta) \\
&\quad \times \int \mathrm{d}^3\mathbf{A}\,\mathrm{d}^3\mathbf{B}\, |\mathbf{A} \wedge \mathbf{B}| \rho(\mathbf{A}, \mathbf{B}; \mathbf{p}, \mathbf{q}). \quad (4.6.31)
\end{aligned}
$$

Since $\mathbf{p}, \mathbf{q}$ are now both parallel in the z-direction, F in (4.6.30) can be found using (4.6.10):

$$
F = \frac{k_2(p^2 + q^2)}{30}[(s_y^2 + t_x^2 - s_x t_y - s_y t_x) + 2(s_x^2 + s_z^2 + t_y^2 + t_z^2)]. \quad (4.6.32)
$$

Since F is now a quadratic form in the components of $\mathbf{s}, \mathbf{t}$, they are straightforward gaussian integrals. The method of calculation follows exactly that of the previous section, only the final integral with respect to the vector $\mathbf{t}$ (from Fourier transforming the modulus of the vector product) is particularly difficult, is integrated numerically to give

$$
d_{\mathrm{L},3} = 0.21360 k_2, \quad (4.6.33)
$$

which is very close to the C line density (4.6.27).

The density of disclination points in three dimensions is another case of equation (6.25) of [Hal81], and is simply

$$
d_{\mathrm{disc},3} = \frac{1}{\pi^2}\left(\frac{k_2}{6\pi}\right)^{3/2}. \quad (4.6.34)
$$

Singularity type	General value $/k_2$	Shell spectrum (in appropriate units of λ_{m})	Planck spectrum (in appropriate units of λ_T)
Dislocation density $d_{D,3}$	0.106103	4.18879	78.7461
C line density $d_{\mathrm{C},3}$	0.21096	8.3283	156.57
L line density $d_{\mathrm{L},3}$	0.21360	8.4326	158.53
Disclination density $d_{\mathrm{disc},\,3}$	0.001238	0.3071	25.03

Table 4.2: Numerical statistical densities of polarization singularities and comparison with spatial dislocation density, for general values, the shell spectrum and the Planck spectrum.

4.6.4 Summary of statistical densities of polarization singularities in three dimensions

The numerical results of the foregoing calculations are summarised in table (4.2), as well as the values for the shell and Planck spectra.

The C, L line densities are numerically about twice the dislocation density. In the paraxial case, the C point density was exactly twice the dislocation density, since paraxial C points are dislocations of the right or left hand circular components of the vector. However, there appears to be no simple explanation for the factor of almost 2 ($d_{\mathrm{C},3}/d_{D,3} = 1.988$). It is possible that the factor 2 is related the index 1/2 nature of C lines, and the field φ in which the field components are quadratic has phase singularities where the C lines are. It is not clear why the L lines should have almost equal density to the C lines.

Following the discussion in section 3.3.1, the density of points where the singularity lines cross a plane is half of the corresponding spatial density. For C lines this is very close to that calculated paraxially, although for C lines crossing a plane, $\mathbf{n}$ is not necessarily perpendicular to this plane.

4.7 Discussion and Conclusions

In this chapter, we have discussed the natural polarization singularities in vector wave-fields, which are geometrically the singularities of polarization ellipses. These are therefore the singularities of light fields when polarization plays a role, and are also the singularities of the electric or magnetic field (in free space). We have indicated that a possible

singularity structure involving the two fields is $\mathbf{E} + \mathrm{i}\mathbf{H}$ (equation (4.4.2)), although this too has problems (the position of the singularities is time dependent, even in monochromatic fields), and we make no claim that this is the final framework for singularities in electromagnetic waves.

By comparison with dislocations, the study experimentally of polarization singularities in optical waves is still in its infancy, and it is possible that future experimental observations will lead to new theoretical understanding.

The local geometric structure of polarization singularities is not as well understood as that of phase singularities. A set of questions which naturally arise here are the corresponding in vector fields to the morphologies of the phase singularity in the polarization scalar φ : its anisotropy ellipse, its twist and twirl, etc. The singular vector structure is richer than for scalars, such as the phenomenon of index switching (of both C and L lines).

Scalar fields do have polarization singularities in them as well, those of the gradient $\nabla\psi$ (polarization singularities are not prohibited by the fact that this field is curl-free). We have seen that the L singularities in $\nabla\psi$ are important in dislocation reconnection, as shall be developed in the next chapter. Can more dislocation geometry be described using polarization singularities? We shall see in the next chapter that C lines appear for twirling loops, but there is more still to be understood.

Although the formalism of the earlier chapters is new in places, the real significance of this work is in the statistical calculations of singularity densities in sections 4.5, 4.6. With the methods employed in this thesis, however, it does not seem to be possible to go further than just the density calculations, which were difficult enough. The richer vector structure means that there are more possible quantities to average; as well as those of the last chapter (speed, curvature, etc), the density of index switching points is another obvious calculation that may be made (unfortunately, these integrals involve quantities in the exponent with powers higher than quadratic, and the usual gaussian integration methods cannot be applied).

Chapter 5

The topology of twisted wavefronts and knotted nothings

> My soul is an entangled knot
> Upon a liquid vortex wrought
> By Intellect, in the unseen residing,
> And thine doth like a convict sit,
> With marlinspike untwisting it,
> Only to find its knottiness abiding;

J. C. Maxwell, *A Paradoxical Ode*, 1878, in L. Campbell and W. Garnett, *The life of James Clerk Maxwell*, Macmillan, London, 1882

So far, in three dimensions, only the local structure of dislocation lines has been considered, and in this chapter we shall explore global properties of dislocation lines, particularly the case of closed loops. Using the notion of twist introduced and discussed in section 2.8, we shall see how dislocation loops give rise to nontrivial and interesting topology, both of the phase surfaces (wavefronts), and in their knotting and linking. The first section 5.1 is a general discussion, concerned only with the topology of closed nodal lines in complex scalar wavefields, not involving any wave equation directly (as with most of the material of chapter 2). Much work concerning the topology of closed phase singularity loops has been done by Winfree and collaborators [WS83a, WS83b, WS83c, WS84, WWS85, Win87, Win95]; the relevant mathematical theory is that of fibred spaces [ST80, Sei80].

The remainder of the chapter is concerned with the details of a particular construc-

tion of torus knots and links in the nodal lines of solutions of the Helmholtz equation, introduced in [BD01a], and explained in section 5.2. The construction is applied to Bessel beams in section 5.3, to polynomial waves in 5.4, and to paraxial waves (both paraxial polynomial and Laguerre-Gauss beams) in 5.5, where the construction is illustrated by explicit calculations for the trefoil knot and Hopf (simple) link. After the conclusions, there are two appendical sections, 5.7, where the different types of wave beam are defined, and 5.8, which explains the (not insurmountable) difficulties with the knot construction for paraxial waves. Unlike vortex knots in fluid dynamics [Kel67, Kel69, Mof69, RSB99], dislocation knots can be untied, by the mechanism of reconnection explained in section 2.6; the dissolution of polynomial knots is discussed in sections 5.4, 5.5.

5.1   Twisted loops and dislocation threading

The crucial result of this section is what we call the 'twisted loop' theorem, discussed in various forms and cases by Winfree and collaborators (see references above). It was rediscovered in the present context by Dennis [Den00] in an attempt to understand the topological nature of screw dislocations (such as (2.8.1)): in particular, is it possible for such a dislocation, with a uniform phase twist, to be a closed loop? The problem may be understood more clearly by considering the phase surface for any phase χ mod π in the neighbourhood of the singular line, which shall be referred to as the χ-ribbon (the dislocation is on the axis of this ribbon, as in figure (5.1)).

For a closed dislocation loop, each χ-ribbon must be closed, and since the field is continuous, the total number of 2π twists about the ribbon axis must be a signed integer. The sign of this *screw number* agrees with the sign of the phase helicoid twist defined in (2.8.2), negative for a left-handed helicoid, positive for a right-handed helicoid. Topologically, it is the *linking number* of the edges of the ribbon (see, for example [Ada94] page 18), with directions given by the direction of topological current for the dislocation on the ribbon axis. This is illustrated in figure (5.1). The number cannot be a half-integer (such as a Möbius band), because the phase surface is a zero contour of the function $u(\chi) = \xi \cos \chi + \eta \sin \chi$, and separates regions in space where the function $u(\chi)$ is positive from regions where it is negative.

The geometry of twisted closed ribbons and their linking number Lk is described by
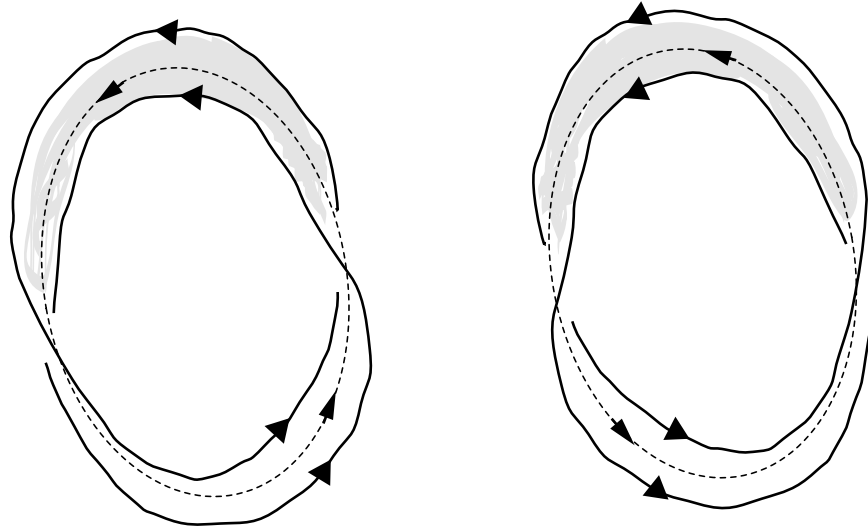
Figure 5.1: Two phase ribbons, with the sign of screw number given by the linking number of the ribbon edges (the dislocation on the axis is a dotted line). The left hand ribbon is twisted in a left-handed sense, and the linking number of the ribbon edges is -1 in this case. The loop on the right is twisted in a right-handed sense, and the linking number of its edges is $+1$.

the Calugareanu-White-Fuller theorem, which states that

$$Lk = Wr + \int Tw, \tag{5.1.1}$$

where $\int Tw$ is the integral over the loop (with respect to arclength) of the local twist (2.8.2), and Wr is the *writhe* of the dislocation curve (this theorem is described in more detail in [BF71, Han98a] and [Ada94] section 7.1). The writhe is a global measure of the departure from planarity of the curve, and we do not consider it further (it is zero if the curve is planar). Both Wr and $\int Tw$ change if the ribbon's configuration is adjusted (*isotopic transformation*), but the linking number Lk is a topological invariant of the curve. Any of the various twist averages $Tw_\chi, Tw_\phi, \sigma, Tw_{tw}$ can be used in (5.1.1), since the appropriate averaging integral (in χ, ϕ, etc) commutes with the arc length integral. All of the phase ribbons must twist the same number of times; otherwise one of the surfaces crosses another on the dislocation, and as it does so the vorticity $\boldsymbol{\omega}$ will vanish. This happens on a dislocation generically at a reconnection, and therefore does not apply to a single loop. Lk, which we shall henceforward call the screw number, is therefore unambiguously defined for the dislocation.

We are now in a position to state the twisted loop theorem (see also [WS83b, WWS85, Win87, BD01a]):

Theorem [The twisted loop theorem] *The screw number m of a closed, strength 1 dislocation loop is equal to minus the dislocation strength threading the loop (in a right hand sense with respect to the loop).*

The proof is as follows: let the dislocation loop be L and consider a closed curve C just inside L, with the same orientation as the dislocation. Around the dislocation, each phase helicoid mod 2π crosses C m times, and the phase at C changes by $-2\pi m$ (the sign coming from the screw sense; consider figure 5.1). This is the dislocation strength threading C, ie

$$\begin{aligned} -m &= \frac{1}{2\pi} \oint_C \mathrm{d}\chi \\ &= \text{dislocation strength threading } C, \text{ by } (1.2.4). \end{aligned} \qquad (5.1.2)$$

QED

This result is remarkable, and supports the philosophy in chapter 1 that singularities organise the phase structure of the field; in this case, the strength of the singularities threading a loop is related to how the phase twists around the loop. The local structure if the singularity is therefore playing a direct role in the field topology.

Much care needs to be taken as to precisely what signs and topologies are really allowed; the discussion in [BD01a] appendix B may appear slightly confusing, since the screw number defined there is actually minus the screw number m defined here as the linking number of the edges of the phase ribbon. Possible confusion increases when one looks at the topology of the phase surfaces (continuing the ribbon out from near the dislocation); [WS83b], figure 7 shows an optical illusion masquerading as a possible phase surface, and give more sources for confusion.

The twisted ribbon nature of the phase contours near the dislocation has consequences for the global topology of the phase surface. With experimentation,[1] the only way one can connect up a twisted, two-sided ribbon is with one half-twisted bridge, as in figure (5.2). Each phase surface of a twisted dislocation must have such a bridge, and all the bridges must cross each other within the loop - this is an alternative proof of the theorem. The

---

[1]I used the 'poor topologist's standby' of paper and plasticine for topological experiments; Winfree and Strogatz [WS83a, WS83b] used dental wax.
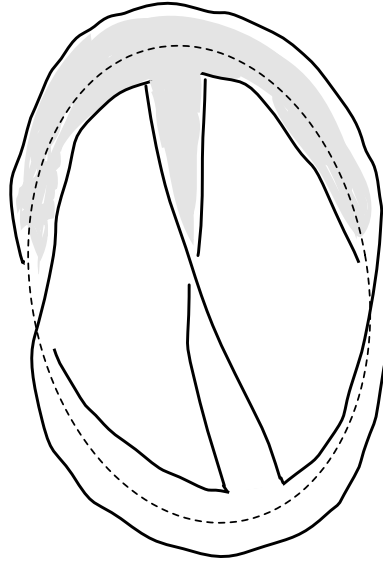
Figure 5.2: The two-sided twisted ribbon is 'closed up' with a half-twisted bridge. This surface is topologically equivalent (homeomorphic) to a torus with a point removed.

resulting surface, with a bit of visualisation or experimentation, is topologically equivalent to a surface of genus 1, that can be embedded nontrivially on a torus (ie a surface of genus at least 1). In fact, as discussed by [Ada94], a twisted ribbon cannot be embedded on a sphere, but only on a torus or other high-genus surface. The reader unfamiliar with topology can find more details of these concepts and terms in [Ada94, FG82].

The twisted loop theorem can be generalised to incorporate the possibility that the screw loop may be (nongenerically) of strength higher than 1, say n. The screw number m is the general case by (5.1.2), equal to the dislocation strength threading the loop; now each phase contour is n ribbons crossing at the dislocation (as discussed in section 2.3), which twists $2\pi m/n$ times around the loop. The twisted loop theorem for high-strength loops is the main implement for making the knot structures of section 5.2. In [BD01a], the screw number is defined differently, as the number of rotations of the dislocation core around the loop, and is quantised in units of $1/n$.

The threading dislocations themselves may be closed loops. Assuming that there is only one (possibly degenerate) dislocation A threading the loop L, if A is a closed loop, then A, L are linked, and A has a screw number $-n$ equal to minus the strength of L : each loop's strength is equal to the other's screw number.

A further generalisation of the twisted loop theorem is that the loop L may be knotted

(and not necessarily threaded), the consequences for which are discussed by [WS83c]. Any knot can mathematically be constructed from several unknotted loops that undergo reconnections of the type described in section 2.6. An analogous construction may be used to make knotted vortices in fluid dynamics [Mof69, MR92], and is topologically equivalent to the Seifert algorithm [Sei80, Ada94], which constructs a surface with the knot as boundary. For phase singularities, the knot is effectively threading itself, inducing a screw number by virtue of its knottiness alone, and any wavefront is a Seifert surface for the knot. [WS83c] apply the Calugareanu-White-Fuller theorem on a projection of the trefoil knot to find that the screw number of the trefoil knot is ∓ 3 (negative for a right-handed trefoil knot, positive for left-handed).

We now turn to the wave equation to show that a closed screw dislocated wave is possible. Working in cylindrical coordinates (R, ϕ, z), we wish to construct a wave ψ with a closed circular loop in the $z = 0$ plane, centred at the origin with radius R_0. For now, we wish the loop to have strength 1. We also want a strength m dislocation up the axis $(R = 0)$, and shall assume, for simplicity, that $m > 0$. Throughout this chapter we shall only work with monochromatic waves, where once and for all we set wavenumber $k = 1$. It is easily verified [Den00] that the wave

$$\psi = (R \exp(\mathrm{i}\phi))^m \exp(\mathrm{i}z)((R^2 - R_0^2) + 2\mathrm{i}(m+1)z) \tag{5.1.3}$$

satisfies the requirements, and solves the Helmholtz equation (1.5.2). If $m = 0$, there is no twist in the dislocation loop, and any constant ϕ section is the double edge dislocation (2.1.10). Even if $m \neq 0$, the screw loop is edge-like in the sense described before; the overall wave propagation direction is in the z-direction (due to the $\exp(\mathrm{i}z)$ factor in (5.1.3)), perpendicular to the twisted loop. Each of the different twist measures of (2.8) give $-m/R_0$ as the local twist for this dislocation, which becomes $-m$ on integrating by arc length, as expected.

From the discussion above, for $m = 1$, one expects that a phase contour of (5.1.3) should be topologically equivalent to a torus. Figure (5.3) shows such a phase contour from two different perspectives. A simple triangulation argument shows that this surface (truncated as in the figure) is homeomorphic to the surface shown in (5.2).

Since the surface must contain an infinite straight dislocation line up the z-axis, it cannot be a compact surface (such a surface, like the standard torus, is closed with finite area). It is a 'noncompact torus' (just as the plane is a 'noncompact sphere', as utilised by
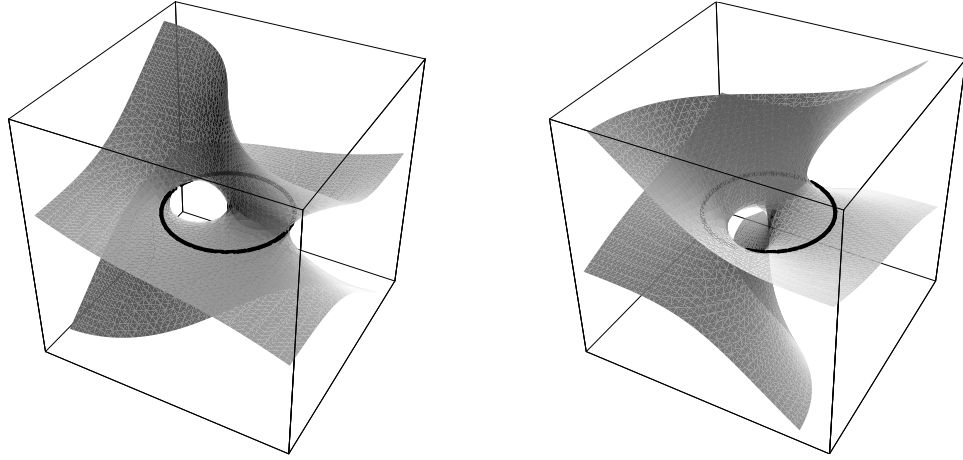
Figure 5.3: A phase surface of the twisted wave (5.1.3) near the closed loop (the black line), shown from two different directions.

stereographic projection, with a point 'at ∞'). Note that, unlike a compact torus, there is no way of distinguishing the two sides of the surface (5.3) (that is, the inside and outside of the surface, where $\operatorname{Re}\psi\exp(-\mathrm{i}\chi)$ is positive and negative, are on an equal footing). Figure (5.4) shows the phase surface (5.3) with its phase conjugate, which is the same surface rotated by $\pi/2$ about the axis.

It has not been possible, using solutions of the wave equation, to close the infinite dislocation line to a loop to make two linked rings; it is easy to construct complex functions with linked zero lines by multiplying together two functions, each containing one of the required nodal lines (this was done by [Fre00], who shows one compact toroidal phase surface in figure 7, but does not observe that its conjugate is noncompact). This non-wave construction can be used to construct more elaborate configurations, such as borromean rings [CBR98], but our attention shall be limited to what is possible with the wave equation.

With appropriate choice of R_0, such as $R_0 = 1$, the core of the looped dislocation in (5.1.3) is elliptical, with major semiaxis in the R-direction, minor in the z-direction. If $m \neq 0$, there is a net phase twirl tw_χ (defined in equation (2.8.10)); the phase lines twist m times around the core ellipse. The geometric twirl tw_ϕ is zero here, as the ellipse does not rotate with respect to parallel transport along the dislocation. It is readily verified that there are $2m$ (degenerate) C lines in $\nabla\psi$ up the z-axis, coinciding with the strength m threading dislocation. This is to be expected: around the looped dislocation, the phase

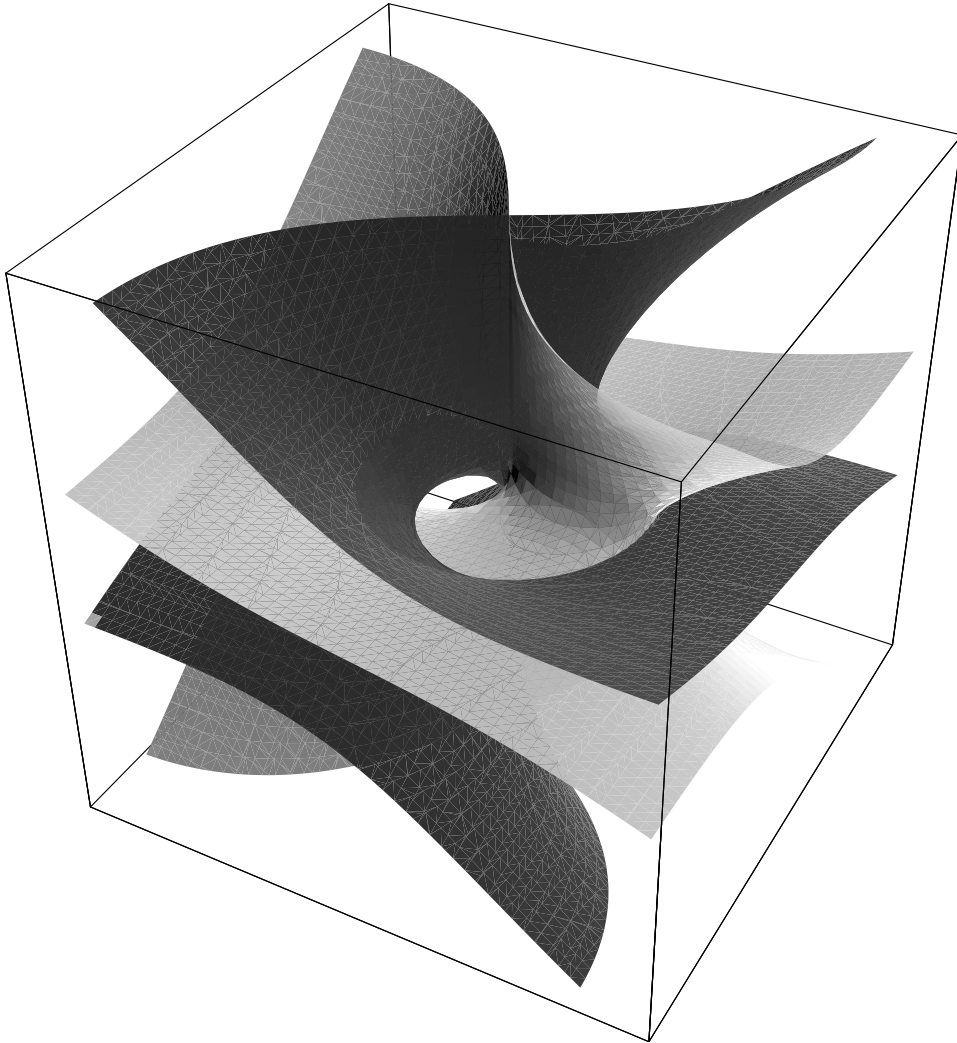Figure 5.4: Two conjugate phase surfaces of the twisted wave (5.1.3) near the closed loop, intersecting on the twisted dislocation lop, and also up the axis.

of the polarization scalar $\varphi \equiv \nabla \psi \cdot \nabla \psi$ changes by 4π. If $m = 0$, there is a single L line threading the loop (we saw this in sections 2.5, 2.6). We shall not investigate the intriguing implications of this here, but observe that a polarization scalar for the polarization scalar may be defined by $\nabla \varphi \cdot \nabla \varphi$, which describes the topological twirl of the C lines in $\nabla \psi$; there is a hierarchy of singularities in the field, organizing the topological structure of closed singular loops in the field.

5.2 A construction for knotted dislocations in the wave equation

The remainder of this chapter shall be concerned with a particular construction, explained fully in [BD01a], which generalises (5.1.3), and in which knotted and linked dislocations exist. Not every knot or link is possible, only the class of *torus knots* and *torus links*, for which the trefoil knot and Hopf link (simple link) are the simplest nontrivial examples; they are knots and links that can be embedded, without crossing, on a torus. Knots such as the figure-8 knot and links such as the borromean rings are not torus knots [Ada94], and are not possible with this construction. Also, the knots and links must be threaded, and so do not have the characteristic screw number for the knot, as discussed in section 5.1. The term knot shall be used to refer either to a knot or link. Unless otherwise stated, cylindrical coordinates (R, ϕ, z) are used, and all waves are monochromatic with wavenumber $k = 1$.

The degenerate structure to be constructed is shown in figure (5.5), consisting of a strength m dislocation A_m along the z-axis ($R = 0$) and a circular loop L_n of strength n with radius R_0 in the $z = 0$ plane, centred on the origin. For convenience, it is assumed that $m, n > 0$, and appropriate generalisation for negative values is obvious. The topological current of L_m is taken to be left-handed with respect to the sense of A_m, as in figure (5.5). This ensures that the topological charge of L_n in the (R, z) plane for $\phi = 0$ is positive, and implies that the helicoid structure local to L_n is right-handed. The wave (5.1.3) is a particular example for $n = 1$. The screw number of L_n is therefore $-m$, and the core rotates $2\pi m/n$ around L_n in a right-handed sense, in the direction of the topological current of L_n. Both A_m and L_n are degenerate structures, sensitive to perturbation, under which they unfold: A_m to an m-stranded helix, L_n to a knotted or linked configuration, as shall be described.
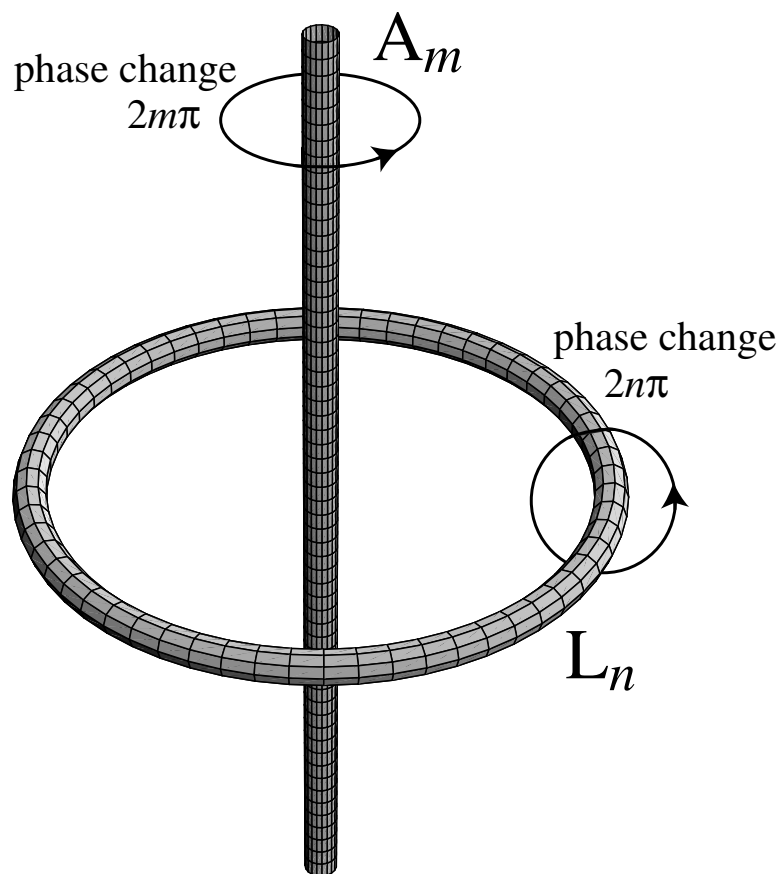
Figure 5.5: Unstable strength m axial dislocation threading unstable strength n dislocation loop L_n (in particular, note the direction of topological current in L_n).

An unperturbed wave ψ_0 with dislocations A_m, L_n must have the following local structure, as required by section 2.3. Near A_m, where K is a (complex) constant,

$$\psi_0 \approx K R^m \exp(im\phi). \tag{5.2.1}$$

Near L_n, where $K_\pm$ are complex constants with $|K_+| > |K_-|$, the local structure is (cf (2.3.7), (5.1.3))

$$\psi_0 \approx (K_+(R - R_0 + iz)^n + K_-(R - R_0 - iz)^n)\exp(im\phi). \tag{5.2.2}$$

In (5.1.3), it is easily shown that $K_\pm = ((R + R_0) \pm 2(m+1))/2$, which is constant when $R \approx R_0$. With these two restrictions, the resulting unperturbed wave has the form

$$\psi_0(\mathbf{r}) = f(R, z)\exp(im\phi), \tag{5.2.3}$$

and L_m can be constructed by imposing conditions on the ϕ-independent function f, and all further ϕ-dependence shall be omitted unless necessary. From (5.2.1), for A_m to be present, ψ_0 must satisfy the $m + 1$ conditions

$$\partial_R^p \psi_0(0, z) = 0 \qquad 0 \le p \le m - 1,$$
$$\partial_R^m \psi_0(0, z) \ne 0, \tag{5.2.4}$$

and from (5.2.2), the $n(n+1)/2$ conditions

$$\partial_R^q \partial_z^{p-q} \psi_0(R_0, 0) = 0, \quad 0 \le q \le p, \quad 0 \le p \le n - 1$$
$$\partial_R^q \partial_z^{n-q} \psi_0(R_0, 0) \ne 0, \quad 0 \le q \le n. \tag{5.2.5}$$

There are $n(n+1)/2$ conditions on L_n, rather than the $n-1$ conditions one might naively assume, because not only do we require, in the (R, z) plane, the coalescence of n strength 1 dislocations, but also $n - 1$ phase saddles, to keep the Poincaré index of the degenerate point 1 (this was described in section 2.3). There are not as many explicit conditions on A_m as L_n because some of the conditions are hidden by the circular symmetry of the cylindrical coordinate system (φ is singular when $R = 0$). In the (R, z) plane, around a positive circuit enclosing the point $(R_0, 0)$, the phase changes by $2\pi n$, so for each phase χ (mod 2π) the phase star has n phase lines emerging from it. By the generalised twisted loop theorem, the star rotates m/n times as ϕ increases from 0 to 2π (in the opposite direction to the topological current of L_n). We now perturb ψ_0 by adding a weak wave
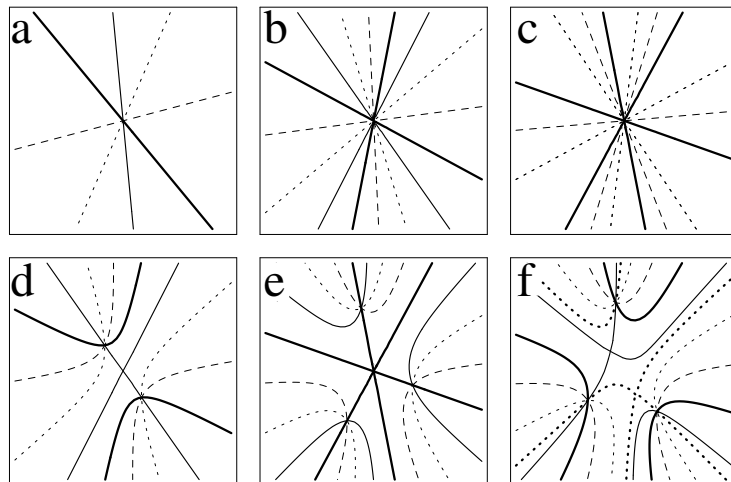
Figure 5.6: Phase contours modulo 2π, in the (R, z) plane. (a),(b),(c) are for a section of unfolded L_1, L_2, L_3, (d) stable unfolded strength 2, (e) unstable unfolded strength 3, (f) stable unfolded strength 3. Note in particular the saddle in (d), two of whose phase lines separate the two dislocations, the phase lines joining them.

ψ_p, which does not contain any dislocations itself in the region of interest (such as a plane wave travelling in the z-direction). The total wave is now (where ε is a small positive numerical parameter)

$$\psi(\mathbf{r}) = \psi_0(\mathbf{r}) + \varepsilon\psi_p(\mathbf{r}). \tag{5.2.6}$$

Upon perturbation, the axial dislocation A_m unfolds to an m-stranded helix. The unfolding for L_n is more complicated. In the (R, z) plane, the degenerate L_n point explodes into n dislocation points and $n - 1$ saddles (which may remain degenerate if the perturbation ψ_p has certain symmetries, as in figure (5.6)). Before perturbation, the phase star in this plane rotated m/n times, and after perturbation the dislocations are n strands, which (as is justified below) rotate in the same way that L_n did.

Since the perturbing wave is small, the pattern of phase lines (mod 2π) in the (R, z) plane, far enough from the point $(R_0, 0)$, is unchanged after the perturbation (since values far enough from the zero are unaffected by the small $\varepsilon\psi_p$). Now, since each of the unfolded dislocations has the same sign, there must be a particular phase line separating the two dislocations and not ending on either of them. The existence of such a contour is clear from continuity arguments, as is the fact that one of the saddles in the unfolding lies on this phase line (the structure is that of figure (5.6d)). Consider the phase $\chi(\phi)$ of one of

these saddles, as ϕ increases (as with the dislocations, the saddle locus line may not join up with itself in a 2π circuit of ϕ). If the phase of $\chi(\phi)$ does not change by a multiple of 2π, then it is clear that the asymptotic contour must also turn m/n times, with the rest of the asymptotic pattern. It is shown in [BD01a] appendix C that this condition is satisfied provided ψ_p does not itself contain dislocations threading L_n. The dislocation phase stars, being separated by the saddle contour, must themselves turn m/n times. The 'saddle paddle' thus convects the unfolded dislocation pattern, which rotates by $2\pi m/n$ in an azimuthal circuit. If m and n are coprime, the m unfolded dislocation strands cannot be m separate loops, because each dislocation has rotated a noninteger number of times.

It is not difficult to see from the above that the unfolded strands of L_n form an (m,n) torus knot [Ada94], the simplest of which are the (3,2) trefoil knot (whose handedness is the product of signs of m and n). If $n = 1$ as in (5.1.3), the perturbed dislocation is a closed helix. If m and n share a common factor N but are not equal, so that $(m,n) = N(m_0, n_0)$, then L_n unfolds to N identical linked loops, each of which is an (m_0, n_0) torus knot. The simplest such link is therefore the $(2,2)$ torus knot, mathematically known as the Hopf link. The simplest linked knots are the (6,4) torus knot, corresponding to two linked trefoils. The unfolded structures of the torus knot and Hopf link threaded by helices is shown in figure (5.7).

Note that, although the construction gives torus knots, the 'torus' is not a phase surface, and as discussed in section 5.1, a knot's phase surface is usually its Seifert surface. Moreover, since the A_m helix threads the knot itself, the screw number σ of the threaded knot is the knot's characteristic screw number κ minus the dislocation strength threading the loop, in this case $\kappa - \sigma$, which is 0 for the trefoil here.

If ψ_p is cylindrically symmetric (independent of ϕ) then, for small ε, the zeros of $\psi(\mathbf{r})$ of (5.2.6) can be found explicitly. Firstly, in the neighbourhood of L_n, ψ_p is a constant, ie

$$B_L \equiv \psi_p(R_0, 0). \tag{5.2.7}$$

We shall work in polar coordinates (ρ, γ) in the (R, z) plane, defined by

$$R - R_0 + \mathrm{i}z = \rho \exp \mathrm{i}\gamma. \tag{5.2.8}$$

Therefore, from (5.2.2), (5.2.6), the dislocations satisfy

$$K_+\rho^n \exp(\mathrm{i}n\gamma) + K_-\rho^n \exp(-\mathrm{i}n\gamma) = -\varepsilon B_L \exp(-\mathrm{i}m\phi). \tag{5.2.9}$$

a   b
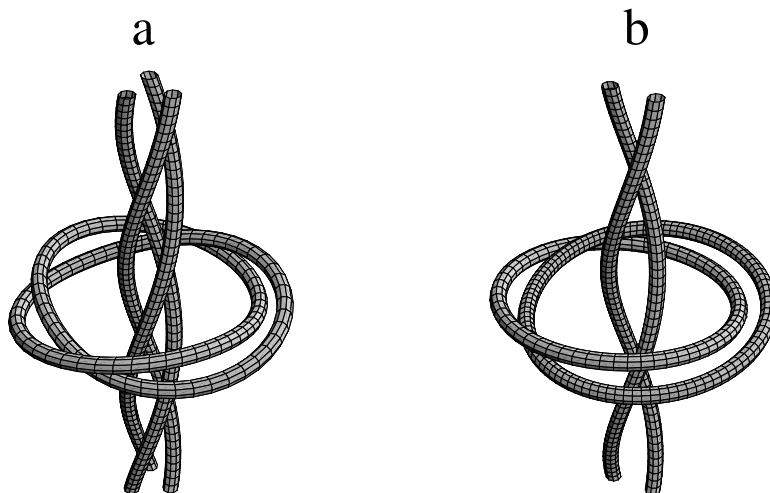
Figure 5.7: Stable unfoldings of figure (5.5): (a) the trefoil knot $(m, n) = (3, 2)$; (b) the Hopf link $(m, n) = (2, 2)$.

As ϕ increases by 2π, the entire argument changes by $-2\pi m$, so (recalling that $|K_+| > |K_-|$) γ, and indeed, the entire pattern, rotates through $-2\pi n/m$; each dislocation strand rotates in a left-handed sense as ϕ increases, by n/m turns. In n azimuthal circuits (ϕ changing by $2\pi n$) the dislocation strand matches with its starting point, for the first time if m, n are coprime, confirming the torus knot structure (with obvious extension if m, n are not coprime). On each azimuthal section, the dislocation points lie on a circle of radius $\rho = \mathcal{O}(\varepsilon^{1/n})$, and the union of all these circles (for each ϕ) gives the torus on which the knot is wound, the coordinates on the torus given by angles ϕ, γ.

In the neighbourhood of A_m, the perturbing wave takes on the form

$$B_A(z) \equiv \psi_p(0, z). \tag{5.2.10}$$

Since ψ_p must be a solution of the wave equation, it must vary with z if it is to be independent of ϕ; for instance, the plane wave $\exp(iz)$ satisfies the wave equation, and has a phase uniformly increasing with z. Using (5.2.1), the m unfolded axial dislocations lie on the axial tube with (variable) radius

$$R(z) = \varepsilon^{1/m} |B_A(z)/K|. \tag{5.2.11}$$

Note that the phase of K is also z-dependent in general. The azimuthal position of the m helical strands are

$$\phi_j(z) = \frac{\arg(B_A(z)/K)}{m} + \frac{2\pi j}{m}, \qquad 1 \le j \le m. \qquad (5.2.12)$$

Since B_A must vary with z, (at least in phase), the strands rotate, in a right handed sense if the argument of B_A increases with respect to z. This is the case in figure (5.7).

If $B_A(z)$ vanishes for certain z values, then A_m unfolds to an m-stranded chain rather than a helix, by (5.2.11); this is the case when one applies the construction here to the electron wavefunctions in atomic hydrogen, as described by [Ber01b].

In the next section, we shall show how this construction can be used in practice, by finding explicit solutions of Bessel beams with nodal lines in the form of the trefoil knot and Hopf link.

5.3   Bessel knots

The remainder of this chapter is concerned with the implementation of the construction in the previous section to various types of wave beams. The conventions and definitions of these wave beams is described in section 5.7. The Bessel beams (5.7.3) are a convenient set of beam solutions, optically realisable experimentally with lasers [Dur87, DMJE87], which satisfy the Helmholtz equation. They possess the required (5.2.1) structure (the (5.2.2) structure is found by choosing an appropriate superposition of the solutions). It is shown in section 5.8 that it is impossible to find a high-strength dislocation transverse to a paraxial beam, which means that there is no way that (5.2.2) can be satisfied for a paraxial solution. However, knots may be constructed in paraxial beams (such as Laguerre-Gauss beams), but this shall not be until section 5.5.

In terms of (5.2.3), each Bessel beam solution (5.7.3), labelled by the order m of the Bessel function and its transverse wavenumber κ, is

$$f_{m\kappa} = J_m(\kappa R) \exp(\mathrm{i}z\sqrt{1-\kappa^2}). \qquad (5.3.1)$$

The transverse wavenumber κ cannot be greater than the total wavenumber 1, and $0 \le \kappa \le 1$. The Bessel beams automatically have the correct structure to satisfy (5.2.4), and in order to have the structure (5.2.2), different transverse wavenumbers κ_l must be chosen

such that the sum

$$f(R, z) = \sum_{l=1}^{n(n+1)/2} a_l f_{m\kappa_l}(R, z). \tag{5.3.2}$$

with real constants a_l. For calculational simplicity, we can choose κ_1 to be equal to 1, and set $a_1 = 1$ without loss of generality. A fixed choice of the other κ_l is then made, and the other a_l and R_0 are adjusted until the $n(n+1)/2$ conditions of (5.2.5) are satisfied, creating the desired loop L_n. (Although the nodes of the Bessel beams (5.7.3) are degenerate cylinders at the zeros of J_m, superpositions such as (5.3.2) are sufficiently generic off the axis to have line zeros.) We find values of κ_l, a_l explicitly for the (3,2) trefoil knot and the (2,2) Hopf link. Since $n = 2$ in both cases, the loop L_2 requires the superposition of three Bessel functions J_m (m is 3 or 2), satisfying the three conditions (5.2.5), and making two dislocations and a saddle coalesce in an azimuthal section. The radial wavevector components κ_l are chosen to be

$$\kappa_1 = 1, \quad \kappa_2 = \frac{1}{3}, \quad \kappa_3 = \frac{2}{3}. \tag{5.3.3}$$

These were chosen so that the zeros of the different $J_m(\kappa_l R)$ were as far from each other as possible, making it easier to find appropriate a_l. For the trefoil, there are now three equations to solve,

$$\begin{aligned}
f(R_0, 0) &= J_3(R_0) + a_2 J_3(R_0/3) + a_3 J_3(2R_0/3) = 0, \\
\partial_R f(R_0, 0) &= J_3'(R_0) + \frac{1}{3} a_2 J_3'(R_0/3) + \frac{2}{3} a_3 J_3'(2R_0/3) = 0, \\
\partial_z f(R_0, 0) &= \frac{\sqrt{8}}{3} a_2 J_3(R_0/3) + \frac{\sqrt{5}}{3} a_3 J_3(2R_0/3) = 0.
\end{aligned} \tag{5.3.4}$$

For the link, the Bessel indices 3 are replaced by 2, but otherwise the three conditions are the same. The three equations (5.3.4) can easily be solved numerically, and for simplicity the lowest two zeros of the Bessel superposition are chosen to coincide; the values of the parameters are

$$\begin{aligned}
a_2 &= 10.0302, \quad a_3 = -3.18960, \quad R_0 = 5.44992 \quad ((3,2) \text{ trefoil knot}) \\
a_2 &= 4.73341, \quad a_3 = -2.70176, \quad R_0 = 4.32636 \quad ((2,2) \text{ Hopf link})
\end{aligned}. \tag{5.3.5}$$

We have therefore constructed functions satisfying the conditions (5.2.4), (5.2.5), to make the degenerate dislocations A_m, L_n.

Since the perturbing beam ψ_p cannot have any dislocations within L_n, the J_0 beam with $\kappa = 1/4$ is chosen, which is independent of ϕ,

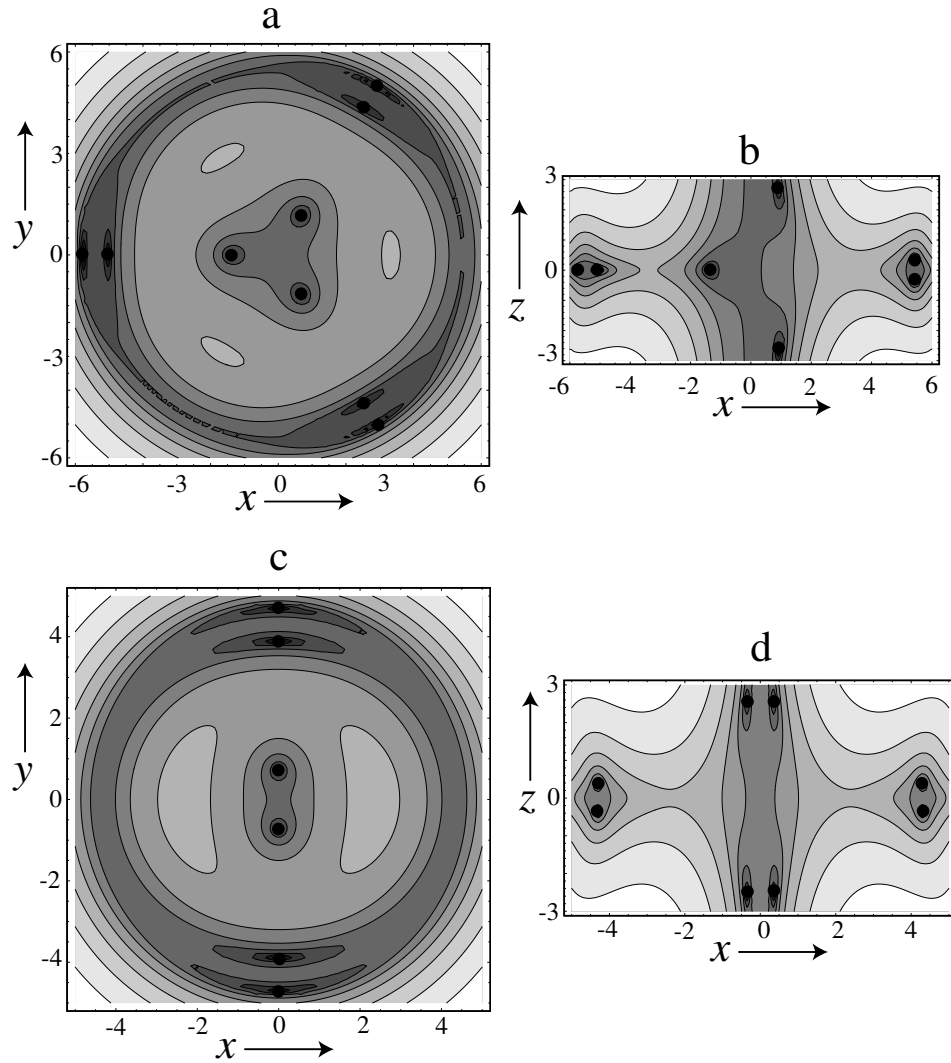$$\psi_p(R, z) = J_0(R/4) \exp(i\sqrt{15}z/4). \tag{5.3.6}$$

Figure 5.8: Density plots of the wave intensity for the Bessel superposition constructed in the text, with $\varepsilon = 0.02$, in the planes $z = 0$ ((a),(c)) and $\phi = 0 \mod \pi$ ((b),(d)). Figures (a),(b) show the (3,2) trefoil knot, (c),(d) the Hopf link. These are sections of the full wave whose knot structures are shown in figure (5.7).

Figure (5.8) shows intensity sections of the total perturbed wave for the two cases. For $R > R_0$, the intensity grows very rapidly, implying that experimental observation of knots using these solutions will be very difficult: a very high sensitivity near the dark places of the beam would be required, so that the entire knot structure does not look like a widening of the dark core, without any contrast picking out the knot structure.

Of course, as we are well aware, light is a vector disturbance, and, especially when paraxial waves are not being used, we cannot assume that the longitudinal component vanishes. Since the vector field must be divergenceless, it is easy to calculate the longitudinal field explicitly. One finds that its zeros are not in the same places as the transverse field. In lasers, therefore, they are not true zeros (although the longitudinal part tends to zero as the wave becomes more paraxial, when the values of $\kappa_l \ll 1$ in (5.3.2)). It makes sense, however, to discuss phase singularities in the transverse component of the beam.

On the subject of experiments, it is perhaps worthwhile to point out that the knot construction (particularly (5.2.1)) requires beams with a factor $\exp(im\phi)$, which are often studied because the phase vortex structure gives rise to orbital angular momentum in the beam (see, for example, [ABSW92, SDAP97, APB99]). This is fortunate, because there is (hopefully) sufficient experimental expertise to physically realise the knot construction. However, it does not seem likely that there is any deeper connection between knots in such beams and their orbital angular momentum, but perhaps a sufficiently imaginative experiment may combine the two properties.

5.4 Knotting and linking in polynomial Helmholtz waves

The construction for knotted and linked dislocations, explained in section 5.2 and implemented for the Bessel beams (solving the Helmholtz beam equation (5.7.1)), rely on constructing the degenerate dislocations A_m, L_n in an initial wave ψ_0, then perturbing them with an undislocated wave ψ_p. Although the knotted configuration is structurally stable, the construction clearly only works provided the perturbation parameter ε in (5.2.6) is sufficiently small. If $R(z)$ in (5.2.11) exceeds R_0, then the helical A_m strands meet the knotting L_n strands, and reconnection events, discussed in section 2.6, take place. ε was called a in [BD01b], and the discussion here follows the arguments of this paper.

The model used to demonstrate this reconnection in action is the family of polynomial Helmholtz waves $g_{\mathrm{H}mt}(R, z)$ (listed in table 5.1). The analogue of (5.3.2) for polynomial

waves is

$$f(R, z) = \sum_{t=0}^{n(n+1)/2} a_t g_{\mathrm{H}mt}(R, z). \tag{5.4.1}$$

As before, the real a_t are chosen such that the degenerate ring L_n is present. For the perturbation, the simplest choice (in beam form) of perturbing wave ψ_p is simply the plane wave in the z-direction

$$\psi_p = 1 = g_{\mathrm{H}00}. \tag{5.4.2}$$

The perturbed beam is therefore (cf (5.2.6))

$$\psi(\mathbf{r}) = \exp(im\phi) f(R, z) + \varepsilon \tag{5.4.3}$$

The perturbation is therefore simply a (changeable) constant.

We are interested in what happens when ε becomes large; in particular, what happens when the strands of the knot reconnect with the threading helix. This occurs, as explained in 2.6, when the dislocations intersect L lines of $\nabla \psi$, which occur where the real vorticity vector field

$$\mathbf{N} = \mathrm{Im}\, \nabla \psi^* \wedge \psi \tag{5.4.4}$$

vanishes ($\mathbf{N} = 2\boldsymbol{\omega}$ here). Since the perturbation (5.4.2) is independent of $\mathbf{r}$, the position of the L lines does not change as ε is varied, and it suffices to find the values of ε for which the dislocations intersect the stationary L line. This L line is stationary only for the beam solution, and is ε-dependent in the proper wave solution, when the beam is multiplied through by $\exp(iz)$. It is used here as a calculational trick to locate the reconnection points; the dislocations have the same positions in the beam and the wave functions.

We first examine the trefoil case, for which, as with the Bessel trefoil, $n = 2$, and t values of 0, 1, 2 are used in (5.4.1). An easy calculation, using the values of $g_{\mathrm{H}mt}$ in table (5.1) for $m = 3$, shows that the choice

$$a_{0,\mathrm{knot}} = 400, \qquad a_{1,\mathrm{knot}} = -40, \qquad a_{2,\mathrm{knot}} = 1, \qquad R_{0,\mathrm{knot}} = \sqrt{20} \tag{5.4.5}$$

satisfies (5.2.5); the overall factor is chosen such that the term with the highest power of R in the superposition (5.4.1) is 1. Explicitly, the unfolded, knotted beam is

$$\psi_{\mathrm{H,knot}}(R, z) = \exp(3i\phi) R^3 \left((R^2 - 20)^2 - 80z^2 + 20iz(R^2 - 20) \right) + \varepsilon. \tag{5.4.6}$$

Figure (5.9a) shows the trefoil when $\varepsilon \approx 0$; note that the unfolded axial strands are at a changeable distance to the axis, which by (5.2.10) is proportional to $1/|5 - z^2 - 5iz|$.

The positions of the L lines can be found by finding the zeros of (5.4.4) with the ψ above, requiring the simultaneous zeros of

$$
\begin{aligned}
N_R &= -480R^5 z \left[3(R^2 - 20)^2 + 160z^2\right], \\
N_\phi &= -40R^5(7R^2 - 60)\left[(R^2 - 20)^2 + 80z^2\right], \\
N_z &= 6R^4 \left[(R^2 - 20)^3(7R^2 - 60) + 1200z^2(R^2 - 20)(R^2 - 12)\right],
\end{aligned} \tag{5.4.7}
$$

which are the loci with (R, z) coordinates

$$
(0, z), \qquad (\sqrt{20}, 0), \qquad \left(\sqrt{\frac{60}{7}}, 0\right). \tag{5.4.8}
$$

On the z axis and the ring $(\sqrt{20}, 0)$, $\psi = \varepsilon$, so the first two L lines correspond to the birth at $\varepsilon = 0$ of the knot and its threading dislocations. On the ring $(\sqrt{60/7}, 0)$, (5.4.6) gives

$$
\psi_{\text{H,knot}}\left(\sqrt{\frac{60}{7}}, \phi, 0\right) = \frac{768000}{343}\sqrt{\frac{15}{7}}\exp(3\mathrm{i}\phi) + \varepsilon. \tag{5.4.9}
$$

The critical value of ε, $\varepsilon_{\text{crit}}$, at which the axial threads meet the knot and reconnect, occurs for

$$
\varepsilon_{\text{crit}} = \frac{768000}{343}\sqrt{\frac{15}{7}} = 3277.66, \qquad R_{\text{crit}} = \sqrt{\frac{60}{7}}, \qquad \phi_{\text{crit}} = \left(\frac{\pi}{3}, \pi, \frac{5\pi}{3}\right), \qquad z_{\text{crit}} = 0. \tag{5.4.10}
$$

Figures (5.9a-d) show the sequence through $\varepsilon_{\text{crit}}$; three simultaneous reconnections (of the type described in section 2.6) take place at $\varepsilon_{\text{crit}}$ (they are not generically simultaneous, as can be demonstrated if, for example, ε depends on cartesian coordinates $\varepsilon = \varepsilon_0 + \varepsilon_x x + \varepsilon_y y$). Note also how, as ε approaches $\varepsilon_{\text{crit}}$, the threading dislocations distort, in preparation for the reconnection which requires the meeting dislocations to be locally coplanar and cotangent.

The construction of the Hopf link is identical to the trefoil case described above, and is explained in detail in [BD01b]. Its parameters in the superposition (5.4.1) are

$$
a_{0,\text{link}} = 144, \qquad a_{1,\text{link}} = -24, \qquad a_{2,\text{link}} = 1, \qquad R_{0,\text{link}} = \sqrt{12}, \tag{5.4.11}
$$

giving the beam

$$
\psi_{\text{H,link}} = \exp(2\mathrm{i}\phi)R^2\left((R^2 - 12) - 48z^2 + 16\mathrm{i}z(R^2 - 12)\right) + \varepsilon. \tag{5.4.12}
$$

which has L lines with (R, z) loci

$$
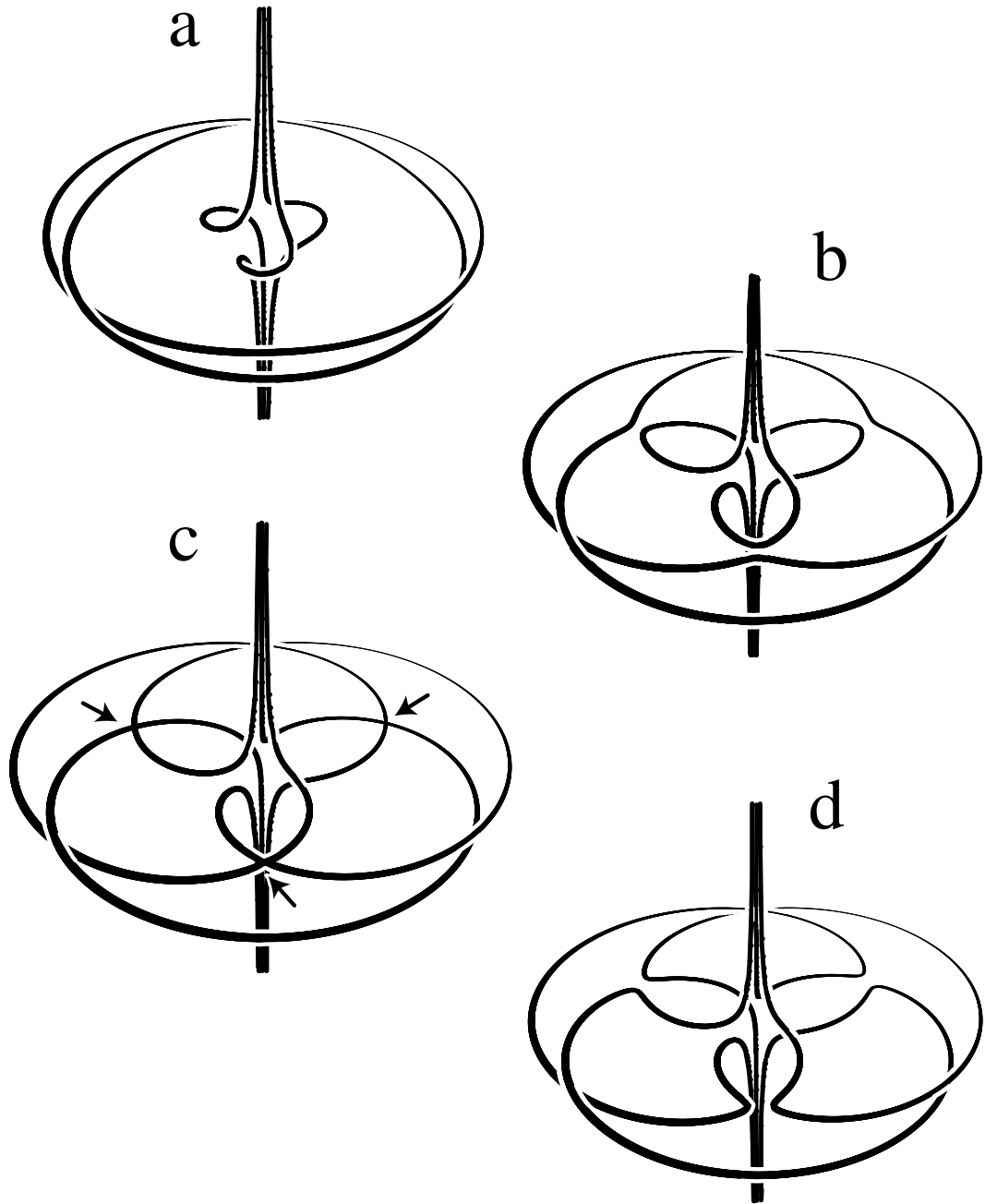(0, z), \qquad (\sqrt{12}, 0), \qquad (2, 0), \tag{5.4.13}
$$

Figure 5.9: Destruction of the trefoil knot in the polynomial Helmholtz wave (5.4.6), as ε increases through (a)-(d). At (c), $\varepsilon = \varepsilon_{\mathrm{crit}}$, (5.4.10), and the three reconnection events are indicated by arrows.

the third of which being the L line at which the two reconnection takes place, with parameter critical values

$$\varepsilon_{\text{crit}} = 256, \qquad R_{\text{crit}} = 2, \qquad \phi_{\text{crit}} = \left(\frac{\pi}{2}, \frac{3\pi}{2}\right), \qquad z_{\text{crit}} = 0. \qquad (5.4.14)$$

The creation and destruction of the link as ε changes is shown in figure (5.10). Note that the values of $\varepsilon_{\text{crit}}$ appear very large; this is an artefact of the values of a_t, R_0 chosen for mathematical simplicity. For instance, if $a_{0,\text{knot}}$ were 1 instead of 400, $\varepsilon_{\text{crit}} = 8.194$, and if R_0 were rescaled to a smaller value $\varepsilon_{\text{crit}}$ would drop further.

5.5   Knots and links in paraxial waves

The reader has already been directed to section 5.8, where it is shown that it is impossible to find a paraxial wave with a loop L_n with $n > 1$. This problem is not insurmountable; although we cannot create the degenerate L_n structure, we can impose slightly less restrictive conditions which still, on unfolding, give torus knots and links, although the change in topology is different from that described in the previous section. For ease of analysis, the polynomial paraxial waves tabulated in table 5.2 shall be used to describe the structure, and the section concludes with appropriate choices of parameter giving the trefoil and Hopf link in Laguerre-Gauss beams.

Although the recovery of the knot construction for paraxial waves was missed at first,[2] it is hardly surprising that knots can still be made: between (5.7.1) and (5.7.4) there is only one term different, $\partial_z^2 \psi$. One can treat this as a perturbing term to interpolate between Helmholtz and paraxial solutions, and if the knot perturbation parameter ε is numerically larger than this, the knot survives, as shall be demonstrated; alternatively stated, 'knot space is bigger than paraxially prohibited space'.

Solutions of the paraxial wave equation automatically have the $\exp(\mathrm{i}z)$ factor removed, so, as with Helmholtz polynomial waves, the knot perturbation can take place with ψ_p being the plane wave in z, which is just the constant 1. By analogy with (5.4.1), the unperturbed wave is a superposition of paraxial polynomial waves $g_{\mathrm{P}mt}$ from table 5.2,

$$f(R, z) = \sum_{t=0}^{n(n+1)/2} a_t g_{\mathrm{P}mt}(R, z), \qquad (5.5.1)$$

---

[2]It was found almost by accident, from a misunderstanding between the theorists (Michael Berry and myself) and experimentalists Johannes Courtial and Miles Padgett, who were trying to implement the Bessel construction experimentally.
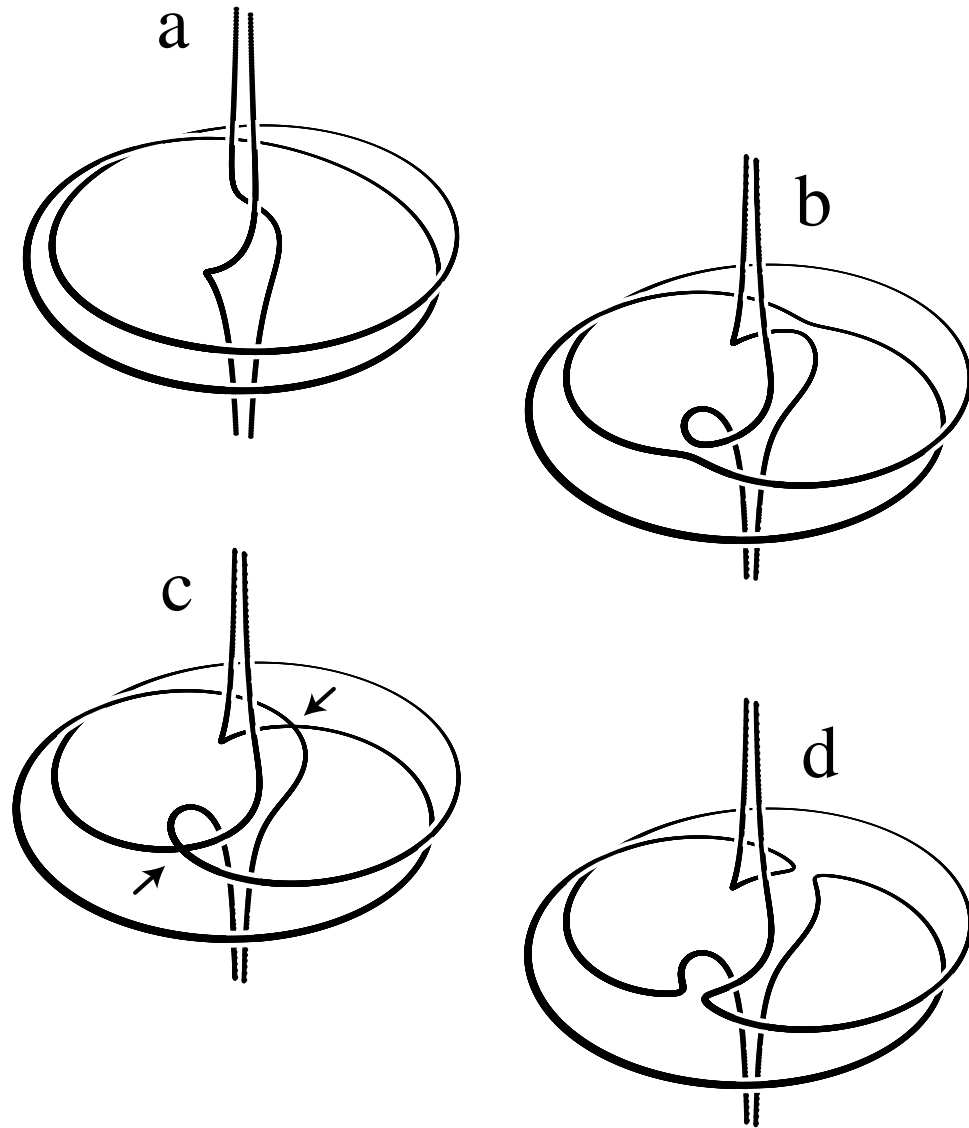
Figure 5.10: Destruction of the Hopf link in the polynomial Helmholtz wave (5.4.12), as ε increases through (a)-(d). At (c), $\varepsilon = \varepsilon_{\mathrm{crit}}$, (5.4.14), and the two reconnection events are indicated by arrows.

which automatically satisfy the A_m conditions (5.2.4), but not the L_n conditions (5.2.5). Because the paraxial case is more complex than the Helmholtz, we shall work with the very simplest knot example, which is the Hopf link.

In this case, (5.5.1) becomes

$$f_{\text{link}}(R, z) = a_0(g_{\text{P20}} + \alpha g_{\text{P21}} + \beta g_{\text{P22}}), \tag{5.5.2}$$

and the L_2 conditions (5.2.5) are

$$f(R_0, 0) = 0, \qquad \partial_R f(R_0, 0) = 0, \qquad \partial_z f(R_0, 0) = 0. \tag{5.5.3}$$

It is impossible for these three conditions to be met simultaneously, but we examine the case when only two (the first and one other) are satisfied. If f and f_R vanish at $(R_0, 0)$ then $\alpha = -2/R_0^2, \beta = \alpha^2/4$. It has an unstable strength 0 dislocation where L_2 ring ought to be at $(R_0, 0)$, and two strength 1 loops at

$$(R, z) = \left(\frac{\sqrt{3}}{2} R_0, \pm \frac{R_0^2}{16\sqrt{3}} \right). \tag{5.5.4}$$

These are separated in z, and we shall see that the twist structure of the beam is sufficient for reconnection processes to make two linked rings from these.

The other possibility ($f, f_z = 0$ at $(R_0, 0)$) has two loops, a degenerate strength 1 loop coalesced with a saddle at $(R_0, 0)$, and another strength 1 loop at $(\sqrt{5/3}R_0, 0)$. These are separated in R but not z, and they do not make a link with $\psi_p = 1$ (although it might be possible with a different perturbing function).

Therefore we choose $a_0 = 1/\beta, \alpha = -2/R_0^2, \beta = \alpha^2/4$, and the resulting unperturbed wave is

$$f(R, z) = a_0 R^2((R^2 - R_0^2) - 48z^2 + 4iz(4R^2 - R_0^2)). \tag{5.5.5}$$

Since only two out of the three conditions (5.5.3) are satisfied, but we added three waves, there is an overall degree of freedom (expressed by the fact that R_0 is not fixed in (5.5.5), although R and z scale differently with respect to R_0). For reason of comparison with the Helmholtz case, a value of $R_0 = \sqrt{12}$ is chosen. The perturbed wave is therefore (cf (5.4.12))

$$\psi_{\text{P,link}} = \exp(2i\phi)R^2((R^2 - 12) - 48z^2 + 16iz(R^2 - 9)) + \varepsilon. \tag{5.5.6}$$

The only difference between $\psi_{\text{P,link}}$ and $\psi_{\text{H,link}}$ is the final numerical coefficient - 9 instead of 12 - but this is sufficient to change the structure from a degenerate strength 2 dislocation

to the two rings at $(3, \pm 1/4)$ and null ring at $(\sqrt{12}, 0)$. As ε increases from 0, the null ring unfolds to two thin crescent-shaped loops lying close to the $z = 0$ plane, as in figure (5.11a).

The reconnection geometry is again provided by the L lines, which once again are independent of ε. They are the zeros of the $\mathbf{N}$ vector field (5.4.4), which has cylindrical components

$$
\begin{aligned}
N_R &= -64R^4 z \left[5R^4 - 72R^2 + 216 + 144z^2 \right], \\
N_\phi &= -192R^3 \left[(R^2 - 4)(R^2 - 9)(R^2 - 12) + 48z^2(R^2 - 3) \right], \\
N_z &= 4R^3 [3(R^2 - 12)^3(R^2 - 4) + 240z^2(72 - 18R^2 + R^4) + 2304z^4]. \quad (5.5.7)
\end{aligned}
$$

The L lines are at the (R, z) positions

$$
(R, z) = (0, z), \qquad (\sqrt{12}, 0), \qquad (2, 0), \qquad (R_{\text{crit1}}, \pm z_{\text{crit1}}). \qquad (5.5.8)
$$

The first three of these are the same as for the paraxial case (5.5.8), respectively the degenerate axial dislocation, the null ring where the crescents appear, and the third to play a similar role as before. There are two further L loops with no Helmholtz counterpart, where

$$
\begin{aligned}
R_{\text{crit1}} &= \sqrt{2 + 4\sqrt{2}\cos\mu} \approx 3.15637, \\
z_{\text{crit1}} &= \frac{1}{6}\sqrt{-43 + 52\sqrt{7}\cos\mu - 70\cos 2\mu} \approx 0.0187059, \\
\text{with } \mu &= \frac{1}{3}\left(\pi - \arccos\frac{47}{28\sqrt{7}} \right). \quad (5.5.9)
\end{aligned}
$$

The link is created when the two strength 1 dislocation rings (5.5.4) meet the two crescents which appeared at the null ring $(\sqrt{12}, 0)$, (four reconnection events) when

$$
\varepsilon_{\text{crit1}} = 37.8161, \qquad \phi_{\text{crit1}} = 65.31° \qquad (5.5.10)
$$

Two of the reconnections involve the upper ring, at $+z_{\text{crit1}}, \phi_{\text{crit1}}$ and $\pi + \phi_{\text{crit1}}$. The other two involve the lower ring, at $-z_{\text{crit1}}, -\phi_{\text{crit1}}$ and $\pi - \phi_{\text{crit1}}$. The sequence of configurations of formation of the link, as ε increases, is shown in figures (5.11a-d). The mechanism is more complicated, but also more general, than the unfolding of L_n, and this method was inadvertently discovered. The specific details, or how this construction may be implemented generally by an analogue of the L_n conditions (5.2.5), have yet to be worked out. The method described can be easily generalised for other cases where $n = 2$ (such
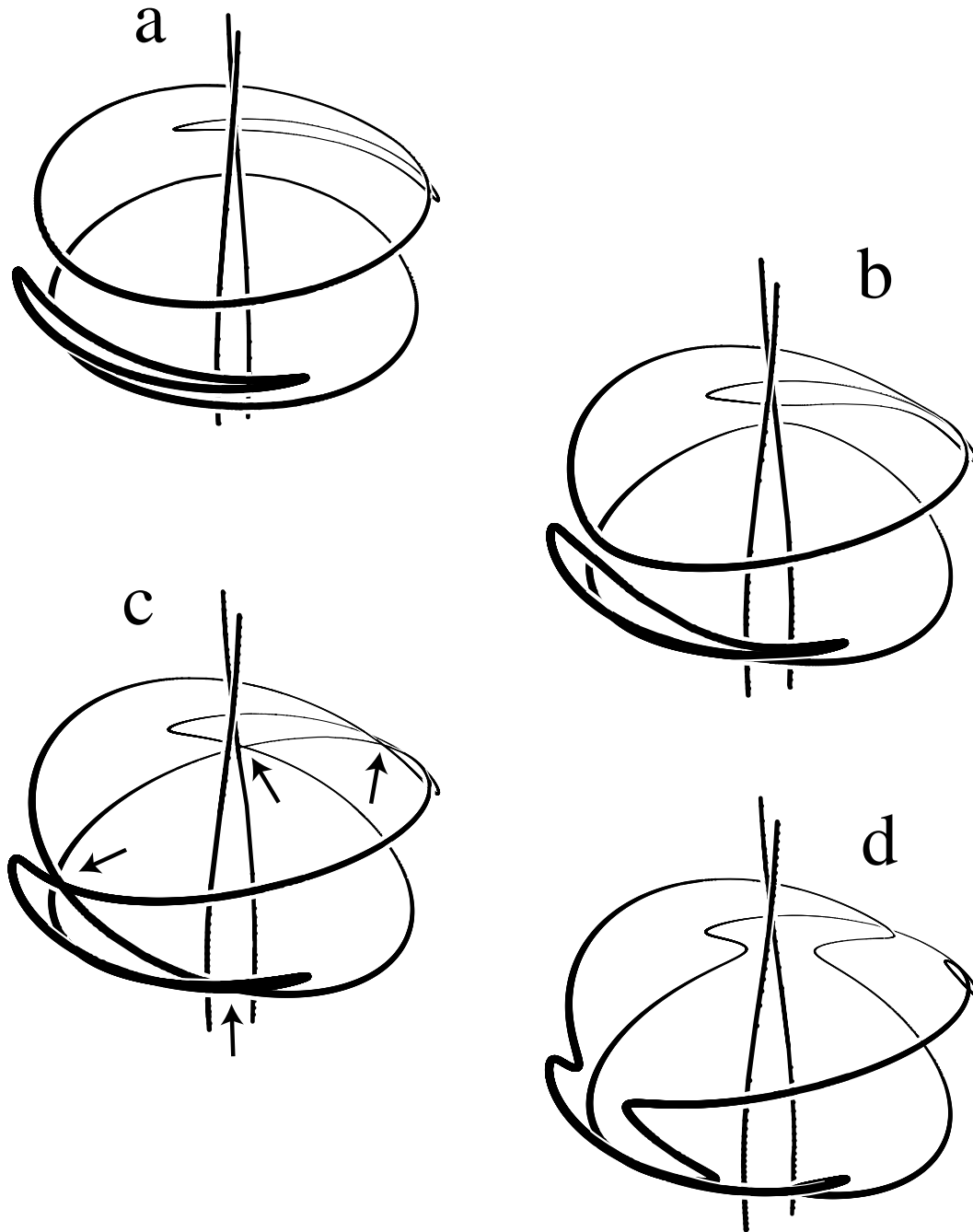
Figure 5.11: Creation of the Hopf link in the polynomial paraxial wave (5.5.6), as ε increases through (a)-(d). In (c), the loop is created by four reconnection events, where $\varepsilon = \varepsilon_{\mathrm{crit1}}$ (5.5.10).
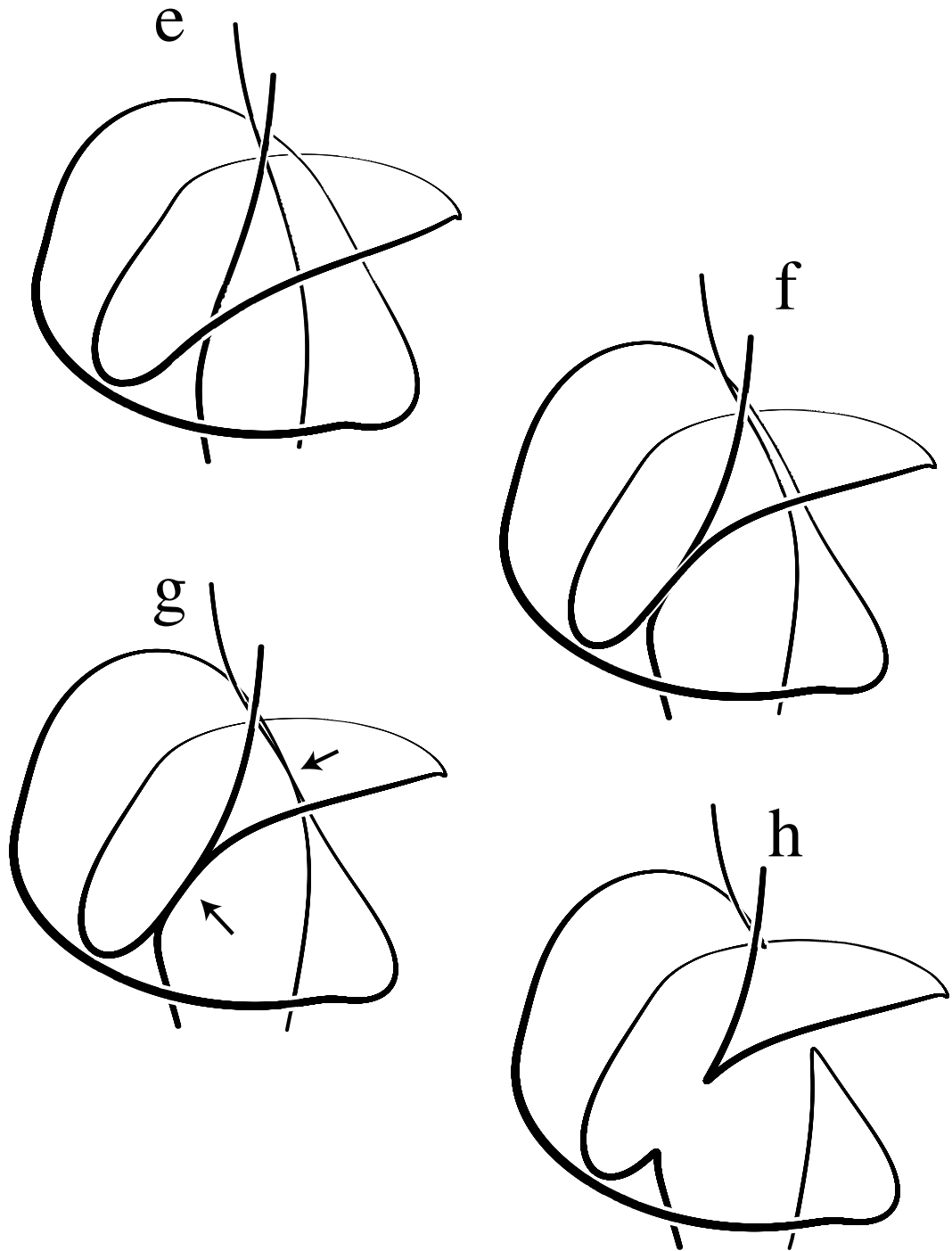
Figure 5.12: The destruction of the Hopf link (5.5.6), continuing the sequence of increasing ε from the previous figure. In (g), there are two reconnection events, where $\varepsilon = \varepsilon_{\text{crit}}$ (5.4.14).

as the trefoil) even when there is no simple immediate paraxial analogue, as shall be seen with Laguerre-Gauss beams. If the paraxial Bessel solutions are used, the trefoil knot and Hopf link are formed by the same process.

After the link has been created, $(\varepsilon > \varepsilon_{\mathrm{crit1}})$, the dislocation topology is the same for paraxial as for Helmholtz waves; the link is eventually destroyed by reconnection with the threading dislocations; the second pair of critical events occurs at the same values of the parameters (5.4.14) as in the Helmholtz case, and are shown in figures (5.12e-h).

For the trefoil knot, the radius R_0 of $\sqrt{20}$ is again chosen to agree with the Helmholtz case, giving the paraxial analogue of (5.4.6),

$$\psi_{\mathrm{P,knot}} = \exp(3\mathrm{i}\phi)R^3[(R^2 - 20)^2 - 80z^2 + 20\mathrm{i}z(R^2 - 16)] + \varepsilon. \tag{5.5.11}$$

The positions of the L lines are the same as in (5.4.8), with an additional two at $(R, z) = (R_{\mathrm{crit1}}, \pm z_{\mathrm{crit1}})$. The critical parameter values of the six reconnections are at

$$
\begin{aligned}
R_{\mathrm{crit1}} &= 4\sqrt{\frac{5(1 + 5\cos\mu)}{21}} \approx 4.160528, \\
z_{\mathrm{crit1}} &= \sqrt{\frac{500(2\cos\mu - \cos 2\mu) - 701}{147}} \approx 0.191897, \\
\text{with } \mu &= \frac{1}{3}\left(\pi - \arccos\frac{2194}{3125}\right).
\end{aligned}
\tag{5.5.12}
$$

Three reconnections take place with the upper ring, at $+z_{\mathrm{crit1}}$, $\phi_{\mathrm{crit1}}, \phi_{\mathrm{crit1}} \pm 2\pi/3$; the others with the lower ring at $-z_{\mathrm{crit1}}$, $-\phi_{\mathrm{crit1}}, \phi_{\mathrm{crit1}} \pm 2\pi/3$. The reconnection with the threading dislocations takes place with the same parameters as before.

To make the link with Laguerre-Gauss beams (5.7.6), one applies the same procedure as above, involving the wave (with perturbation ε),

$$\psi = \psi_{\mathrm{LG20}} - \frac{6}{7}\psi_{\mathrm{LG21}} + \frac{2}{7}\psi_{\mathrm{LG22}} + \varepsilon, \tag{5.5.13}$$

The link is made by exactly the same mechanism as the paraxial polynomial wave. The four reconnection events creating the link (as ε increases) are

$$
\begin{aligned}
\varepsilon_{\mathrm{crit1}} &= 0.003370651, & R_{\mathrm{crit1}} &= 0.883533, \\
z_{\mathrm{crit1}} &= \pm 0.0263448, & \phi_{\mathrm{crit1}} &= \pm 69.41° \text{ and } \mp 110.59°.
\end{aligned}
\tag{5.5.14}
$$

There are four destruction events (since there are two L lines when the threading strands meet the linking strands), at

$$
\begin{aligned}
\varepsilon_{\mathrm{crit2}} &= 0.017968, & R_{\mathrm{crit2}} &= 0.535649, \\
z_{\mathrm{crit2}} &= \pm 0.0388719, & \phi_{\mathrm{crit2}} &= \pm 108.05° \text{ and } \mp 71.95°.
\end{aligned}
\tag{5.5.15}
$$

The knot exists for the wave

$$\psi = \psi_{\mathrm{LG}30} - \frac{8}{13}\psi_{\mathrm{LG}31} + \frac{2}{13}\psi_{\mathrm{LG}32} + \varepsilon, \tag{5.5.16}$$

which is created at the six events with parameters

$$
\begin{aligned}
\varepsilon_{\mathrm{crit1}} &= 0.000972409, & R_{\mathrm{crit1}} &= 0.913606, \\
z_{\mathrm{crit1}} &= \pm 0.0146031, & \phi_{\mathrm{crit1}} &= \mp 74.71°, \pm 45.29°, \pm 165.29°,
\end{aligned} \tag{5.5.17}
$$

and destroyed at the six reconnections where

$$
\begin{aligned}
\varepsilon_{\mathrm{crit2}} &= 0.00573333, & R_{\mathrm{crit1}} &= 0.624702, \\
z_{\mathrm{crit1}} &= \pm 0.0143240, & \phi_{\mathrm{crit1}} &= \mp 53.58°, \pm 66.42°, \pm 186.42°,
\end{aligned} \tag{5.5.18}
$$

The Laguerre-Gauss beams have a different normalisation to the paraxial waves, and the values $\varepsilon_{\mathrm{crit}}$ of the critical events are correspondingly smaller.

5.6 Discussion and Conclusions

We have seen that dislocation loops, in three dimensions, have a rich topology, by virtue of the phase structure they organise. The quantisation of twist has imporant implications, some of which we have explored. However, the corresponding case for twirl has yet to be penetrated; it is likely that this will lead to new understanding of dislocation topology, from the fact that the dislocation is a closed loop in the $\nabla\psi$ field, which encloses C and L lines.

Although we have explored in detail one particular construction for knotted dislocations in the wave equation, we have barely scratched the surface of possibilities of more general dislocation knots. For instance, is there an alternative construction for non-torus knots, such as the figure-8 knot or the borromean rings? Is it possible to find knotted dislocations in solutions to the wave equation where the knots are unthreaded? Also, it is possible that there are connections between knotted dislocations (as current vortices) and knotted vortices in hydrodynamics [Mof69, MR92, RSB99], and possibly the helicity invariant, used the fluid dynamical study of knots, might play a role in knotted dislocation theory.

The knot construction naively applied fails for paraxial waves, but may be recovered. It is not clear what the appropriate generalisation of (5.2.5) should be for a knot with arbitrary n, and further investigation would lead to a deeper understanding of paraxiality.

The constructions here ought to be realisable experimentally, with details for construction of the Hopf link and trefoil knot given explicitly for Bessel and Laguerre-Gauss beams. This is obviously a challenge to experimentalists. The obvious first experimental candidate is an optical one involving lasers, and the beams we have used are easily manipulated in laser optics (of course, at the knot, there may be a small longitudinal component of the electromagnetic vector wave). However, as we have already discussed, there is possibly a problem with these particular constructions, because the intensity grows very quickly with radius R beyond the knot. This is possibly because, for mathematical simplicity, the first two radial zeros of the unperturbed beam were chosen to coincide; the growth of intensity may be less if higher radial zeros were chosen to coincide. Feedback from experimentalists working on a knot realisation could resolve this question.

The knot reconnection studied here is a good example of the general dislocation reconnection described in section 2.6, and the topology can be studied further. As an offshoot of this, as the azimuth ϕ of the plane is varied, dislocation points in the (R, z) plane are annihilated and created (this is especially obvious in figure (5.11)). Preliminary investigation shows some reactions of the form (2.2.5), others of the form (2.2.6). Using the functions here, the relationship between dislocation points in the plane and (curved, twisted) dislocations in space is ripe for further investigation.

As indicated in the discussion to chapter 3, there are a range of questions about the knottedness of dislocations in random waves, whose answer will require different techniques than those used in this thesis to be solved.

5.7   Appendix: Wave beams: Bessel, polynomial and paraxial

The wavefields used in the construction of knots and links are different from the isotropic random waves used elsewhere in this thesis; they are all monochromatic (with wavenumber $k = 1$), and have some form of cylindrical symmetry, possibly with degenerate infinite long straight dislocations up the z-axis. They also all have a factor $\exp(\mathrm{i}k_z z)$ (for appropriate k_z); it is the presence of this factor, along with cylindrical symmetry, that we justify use of the term 'beam' (although beams with cartesian symmetry exist, such as Hermite-Gauss beams, they are not considered here). We shall say that a wave ψ in the form $\psi \exp(-\mathrm{i}z)$ is in *beam form*, and in this section, waves in beam form shall be denoted $\tilde{\psi}$. All of the

waves considered in [BD01b] were in beam form. Solutions of the paraxial wave equation (1.5.5) are naturally in beam form, and it is to compare the subtleties of the different solutions to paraxial and nonparaxial equations that we use this representation.

Since only monochromatic waves are considered, the relevant waves ψ_{H} satisfy the three-dimensional Helmholtz equation (1.5.2) with $k = 1$. Therefore the corresponding beams $\tilde{\psi}_{\mathrm{H}} = \psi_{\mathrm{H}} \exp(-\mathrm{i}z)$ satisfy the equation

$$\nabla_{\perp}^2 \tilde{\psi}_{\mathrm{H}} + \partial_z^2 \tilde{\psi}_{\mathrm{H}} + 2\mathrm{i}\partial_z \tilde{\psi}_{\mathrm{H}}, \tag{5.7.1}$$

where $\nabla_{\perp}^2$ is the transverse laplacian (see (1.5.4)). Cylindrical coordinates are used throughout this section. In beam form, a plane wave travelling in the z-direction is just a constant.

A convenient and realisable set of beams are the set of Bessel beams $\psi_{\mathrm{B},m\kappa}$, which as solutions of the Helmholtz equation are written

$$\psi_{\mathrm{B},m\kappa} = J_m(\kappa R) \exp(\mathrm{i}m\phi) \exp\left(\mathrm{i}z\sqrt{1 - \kappa^2}\right), \tag{5.7.2}$$

where J_m is the Bessel function of first kind, of order $m \geq 0$ [AS65]. As solutions of the beam equation (5.7.1), the Bessel beams have the form

$$\tilde{\psi}_{\mathrm{B},m\kappa} = J_m(\kappa R) \exp(\mathrm{i}m\phi) \exp\left(\mathrm{i}z\left(\sqrt{1 - \kappa^2} - 1\right)\right). \tag{5.7.3}$$

κ is the transverse wavenumber (in units of wavenumber $k = 1$) and obviously (for a nonevanescent wave) $0 \leq \kappa \leq 1$; if $\kappa = 0$, then $\psi_{\mathrm{B},m0}$ is a plane wave if $m = 0$, and zero otherwise. $\kappa = 1$ indicates that the wave propagation is completely transverse. Although used in (5.3.3) for numerical convenience, this is not easily physically realisable; most beams in optics have $\kappa \ll 1$, and in this case the Helmholtz beam equation (5.7.1) becomes the paraxial wave equation (1.5.5) with $k = 1$,

$$\nabla_{\perp}^2 \psi_{\mathrm{P}} + 2\mathrm{i}\partial_z \psi_{\mathrm{P}} = 0; \tag{5.7.4}$$

this is valid in the paraxial approximation. There are subtle differences between non-paraxial and paraxial waves, as the next appendix shows. More understanding of what the paraxial approximation means can be gained by finding the paraxial analogue of the Bessel beams (5.7.3); since κ is small, the factor of z in the exponential approximates to $-\kappa^2/2$, which is exactly the Bessel solution of (5.7.4):

$$J_m(\kappa R) \exp(\mathrm{i}m\phi) \exp(-\mathrm{i}z\kappa^2/2). \tag{5.7.5}$$

As an approximation to (5.7.3), this expression is clearly only valid if κ is small; certainly the beam with $\kappa = 1$ is rather nonparaxial. Note that, if $2\mathrm{i}\partial_z\psi_\mathrm{P}$ is subtracted from each side of (5.7.4), the equation has the same form as the time-dependent Schrödinger equation in the plane (where z plays the role of time t), and all of the paraxial physics of this chapter can be given a quantum interpretation in $2 + 1$ spacetime [BD01b].

Frequently in the study of optical vortices in beams, the set of paraxial solutions known as Laguerre-Gauss modes [ABSW92] are frequently used. They have the form

$$\psi_{\mathrm{LG},mn} = \left(\frac{w(-z)}{w(z)}\right)^n \frac{\exp(-R^2/2w(z))}{w(z)^{m+1}} R^m \exp(\mathrm{i}m\phi) L_n^m\left(\frac{R^2}{|w(z)|^2}\right), \qquad (5.7.6)$$

where L_n^m denotes the associated Laguerre polynomial [AS65], and

$$w(z) = 1 + \mathrm{i}z. \qquad (5.7.7)$$

These formulae apply for beams with waist equal to 1; others are possible by rescaling R and z appropriately. The form of the Laguerre-Gauss beams (5.7.6) given here are slightly different from those given elsewhere (for instance [ABSW92, APB99]), but certainly satisfy (5.7.4).

With dislocations, we are normally only concerned with local properties of waves in the vicinity of the singularity, and global properties of the solution are unimportant. This leads to simpler algebra (for instance the simple forms (1.2.5), (2.8.1) exhibit dislocations without the complications in equations like (5.7.6)). To this end, Nye [Nye98] constructed families of so-called *polynomial waves*, waves that satisfy the wave equation, that in beam form may be polynomials in R and z. They are the polynomial generalisations of the simple linear times exponential waves (like (1.2.5), (2.8.1)) first used to study dislocations [NB74].

We are interested in waves with phase-cylindrical symmetry, where the only ϕ dependence is a factor of $\exp(\mathrm{i}m\phi)$; by inspection, Nye [Nye98] constructed a set of cylindrical solutions to the Helmholtz equation for $m = 0$, which were generalised for any (integer) m by [Den00]. Their use was justified as being local Taylor expansions of waves whose global behaviour would be different (since polynomials diverge for large values of their variables). In fact, they can be constructed exactly by Taylor expanding in κ the Bessel beams (5.7.3). Expanding the Helmholtz beams (5.7.3), one finds in κ^{2n} the term

$$\psi_{\mathrm{H}mn} = R^m \exp(\mathrm{i}m\phi) g_{\mathrm{H}mn}(R, z), \qquad (5.7.8)$$

$n = 0$	1
$n = 1$	$R^2 + 2\mathrm{i}(m+1)z$
$n = 2$	$R^4 - 4(m+2)(m+1)z^2 + 4\mathrm{i}(m+2)z(R^2 - (m+1))$
$n = 3$	$R^6 - 12(m+3)(m+2)z^2(R^2 - 2(m+1)z)$
	$\quad + 2\mathrm{i}(m+3)z(3R^4 - 2(m+2)(3R^2 + 2(m+1)(z^2 - 3)))$

Table 5.1: Polynomials $g_{\mathrm{H}mn}$ from equation (5.7.8), associated with the κ expansion of the Helmholtz Bessel beams (5.7.3).

$n = 0$	1
$n = 1$	$R^2 + 2\mathrm{i}(m+1)z$
$n = 2$	$R^4 - 4(m+2)(m+1)z^2 + 4\mathrm{i}(m+2)zR^2$
$n = 3$	$R^6 - 12(m+3)(m+2)z^2R^2$
	$\quad + 2\mathrm{i}(m+3)z(3R^4 - 4(m+1)(m+2)z^2)$

Table 5.2: Polynomials $g_{\mathrm{P}mn}$, associated with the κ expansion of the paraxial Bessel beams (5.7.5). Note the slight differences with the corresponding polynomials $g_{\mathrm{H}mn}$ in table (5.1).

where $g_{\mathrm{H}mn}(R, z)$ is a polynomial in R and z. The first few $g_{\mathrm{H}mn}$ are listed in table (5.1) (with appropriate choice of overall numerical factors). Similarly, expanding the paraxial beams (5.7.5), one finds polynomials $g_{\mathrm{P}mn}$ in the same way, which are listed in table (5.2).

Expanding in κ is the opposite of the familiar geometrical (or semiclassical) expansion in $1/k$, appropriate in geometrical optics for short waves (large wavenumber). Here, where we are examining the wavelength-level structures of waves, the reciprocal expansion is used, furthering the duality, discussed in the Introduction, between dislocations as singularities in wave optics, and caustics in geometrical optics (whose properties are best studied using geometrical expansions).

5.8 Appendix: The paraxial prohibition against high strength dislocations

The following is a proof of the statement that it is impossible for any solution to the paraxial wave equation (1.5.5), (5.7.4) to have a dislocation with a high strength ($>$ 1) dislocation perpendicular to the propagation direction. It is rare to find restrictions of dislocation morphology in a wave equation, although the high-order structure (2.3.7)

is another example. Physically, the anisotropy between the transverse and longitudinal directions is too much for a high strength dislocation to be possible.

Any loop enclosing the z-axis, of whatever shape, must have at least two points where its direction is perpendicular to the z-axis (in particular, the components of the tangent vector $\mathbf{T}$ must be periodic around the loop, with mean zero, so each must pass through 0 at least twice). Thus paraxial loops of strength n, $|n| > 1$ are not possible if the cartesian paraxial equation

$$\partial_x^2 \psi + \partial_y^2 \psi + 2\mathrm{i}\partial_z \psi = 0 \tag{5.8.1}$$

forbids a strength n dislocation perpendicular to the z-direction.

Coordinates may be chosen so that, at the point of perpendicularity, the dislocation passes through the origin in the y-direction, so the y variation is slower than that in x and z, and the term $\partial_y^2 \psi$ in (5.8.1) is dominated by the other two. In the (x, z) plane, ψ therefore satisfies

$$\partial_x^2 \psi + 2\mathrm{i}\partial_z \psi = 0, \tag{5.8.2}$$

equivalent to the Schrödinger equation in one dimension. The anisotropy between x and z shows that there is no solution proportional to $(x^2 + z^2)^{n/2}$, and, by (5.2.5), (2.3.10), for a strength n dislocation, not only must all derivatives $\partial_x^j \partial_z^{p-j} \psi$ of order $p < n$ vanish, but also must all derivatives $\partial_x^j \partial_z^{n-j} \psi$ *not* vanish. However, (5.8.2) shows that any ∂_z may be replaced by ∂_x^2 (ignoring factors), so that these two sets of conditions may be satisfied simultaneously; if the derivatives of order less than n vanish, then at least one higher derivative with respect to x must vanish too, spoiling the construction.

The prohibition is very subtle, and may be illustrated by the replacement of $\sqrt{1 - \kappa^2} - 1$ with $-\kappa^2/2$ in the Bessel beams (5.7.3), (5.7.5), in the attempt to make a degenerate ring L_2 in the knot construction. One finds that another singularity with strength opposite those in the loop has appeared and combined with them, producing a cancellation, making a degenerate strength 1 object. The paraxial polynomial case was similar, where there was a degenerate strength 0 loop at $(\sqrt{12}, 0)$. These unwanted guests are paraxially inevitable, and even tend to appear for Helmholtz waves as the transverse wavenumber κ decreases, making the numerical solution of sets of equations like (5.3.4) more difficult.

Chapter 6

Singularities in tensor waves: a spinor approach

'When the cube and the things together

Are equal to some discrete number,

[*To solve $x^3 + cx + d$,*]

Find two other numbers differing in this one.

Then you will keep this as a habit

That their product should always be equal

Exactly to the cube of a third of the things.

[*Find u, v such that $u - v = d, uv = (c/3)^3$.*]

The remainder then as a general rule

Of their cube roots subtracted

Will be equal to your principal thing.'

[*Then $x = \sqrt[3]{u} - \sqrt[3]{v}$.*]

Tartaglia's poem explaining the solution of the cubic equation to Cardano [1539], quoted

in Fauvel and Gray, eds, *The History of Mathematics: A Reader*, Macmillan, 1987

In this chapter a formalism is outlined in which the generic polarization singularities of complex vector fields, described in chapter 4, appear naturally as singularities in the geometric description of spin 1 spinors. The generalisations of these spin singularities exist in fields of any spin $s > 1/2$, in particular, spin 2 fields, that may represent gravitational or (tensorial) elastic waves. These generalised spin singularities are still called C lines and

L lines, and are generic in three dimensions with codimension 2, and there are qualitative differences between two and three dimensional fields. The layout of the chapter is as follows: the machinery for looking for spin singularities is described further in section 6.1, spin 1/2 spinor geometry is reviewed in 6.2, and arbitrary spin, and its relation to the Majorana sphere, in section 6.3. The paraxial case is discussed in 6.4, and C,L lines in three dimensional spin fields in sections 6.5, 6.6. The interpretation of spin singularities in vector waves is given in section 6.7, and for spin 2 waves in 6.8. There is an appendical section on the tensor representation of spherical harmonics. The reader is warned that the notation used in this chapter differs significantly in places from that used elsewhere in the thesis.

At a late stage in the writing up of this work, it was realised that there are certain mathematical parallels between the spin geometry explained here (particularly for spin 2, the singularities in tensor wavefields, the understanding of which is the main motivation for this work) and that used in certain classes of solution of the Einstein field equations, the so-called Petrov classification [Syn64]. However, it is unlikely that this subject is familiar to many readers (even those who have got as far as this chapter), and the material is presented more or less as originally intended. The work on this topic is not finished, and is presented in its incomplete state.

6.1   Motivation and introduction

We have discussed at length the geometric properties of phase singularities in complex scalar wavefields, and also saw in chapter 4 that in complex vector waves, the relevant singularities are those of polarization, where the polarization ellipse of the field becomes circular or linear, and the natural orthogonal frame associated with the ellipse is singular.

The generalisation of the notion of polarization and its singularities is possible through the realisation that the fundamental kinds of physical quantity (scalars, vectors, tensors, etc) can be written as representations (via spherical harmonics) of the group of three-dimensional rotations SO(3), and more generally spinors (of integer spin), which are representations of the group of 2×2 special unitary transformations SU(2). Although examples of physical waves discussed here are all of integer spin, we shall develop a general theory of spin singularities (that is, singularities for spinors of spin $s > 1/2$), and only (coherent) states of pure spin s are considered (not mixed states).

Although much of the exposition has a quantum mechanical flavour and formalism (such as the use of the term 'spin' itself), we emphasise that the theory described is general, and the formalism is used because of its familiarity to physicists; much of the mathematics of spinors, tensors and representations grew out of quantum mechanical problems (as in, for instance, [Wig59]).

The primary conceptual object used here to describe polarization of spin s and its singularities is the *Majorana sphere representation* [Maj32], a generalisation of the Riemann sphere for spin 1/2, discussed briefly in section A.3. The Majorana representation is a unique canonical decomposition of a given spin state of spin s into $n \equiv 2s$ unordered spin 1/2 states (that is, unit vectors on the Riemann sphere); thus, up to amplitude and phase, a complex vector (spin 1) is described by two unit vectors, a traceless symmetric complex matrix (spin 2) by four, with rotations R of the state in space corresponding to rigid rotations of the Majorana sphere.

The description appears to have been used first by Majorana in 1932 [Maj32] to describe probabilities of transition between quantum states of atoms in magnetic fields (although its full generality does not appear to have been appreciated in that area [Mec58]). It seems to have been Synge who coined the term 'principal normal directions' in [Syn58] for the directions of the spinors in the Majorana decomposition of the electromagnetic vector $\mathbf{V} = \mathbf{E} + i\mathbf{H}$ (4.4.2), and this was later used by Penrose [Pen60, PR84b] to give a spin interpretation of the Petrov classification [KSMH80] of spin 2 gravitational fields. Penrose later popularised the Majorana picture for quantum states in [Pen89, Pen94], [1] which has earned more of a following in recent years [Leb91, Han98b]. The Majorana sphere formulation is not intrinsically different from the more usual algebraic formulation of spin in terms of eigenfunctions of spin operators and spherical harmonics, but does not appear to admit of a simple description of the addition of angular momentum (facilitated by Clebsch-Gordan coefficients in the algebraic formulation). Geometrically, the Majorana sphere is more appealing, with no axis (such as the z-axis) needing to be chosen, and rotations act directly.

The mathematics of this chapter is more abstract and subtle than that of preceding chapters, and several notational conventions are adopted in this chapter.

- The summation convention, where repeated indices are summed over all values in

---

[1] In these books, he does not mention the significant role played by the Majorana sphere in general relativity, which was one of the reasons why this literature was missed.

that index, is used unless otherwise stated.

- Position in configuration space (the space of the field of spinors) is denoted by the vector $\mathbf{x} = (x_1, x_2)$ (two dimensions) or $\mathbf{x} = (x_1, x_2, x_3)$ (three dimensions), with coordinate indices in lowercase roman i, j, etc. Cartesian tensors are thus written g_{ij}, etc. Only matrices acting on real space (not spin space) are in bold face.

- Spinors in spin space are written in Dirac's bra-ket notation, with $|\psi_s\rangle$ representing a state with spin s (and the subscript is usually omitted). $|\zeta\rangle$ is used when the spin is $1/2$. The spin labels $m = -s, \ldots s$ labelling basis vectors in the spin space are here usually given lowercase greek labels μ, ν, where $\mu = m + s$; the μ represents $0, 1, \ldots, 2s$, and $|\psi\rangle$ is decomposed (employing the summation convention)

$$|\psi_s\rangle = \psi_\mu |\mu\rangle, \tag{6.1.1}$$

where each ψ_μ is a complex number, the μth spin coefficient. When writing vectors in the $2s + 1$-dimensional spin space, square brackets are used, and $|\psi\rangle$ is rewritten $[\psi_{2s}, \ldots, \psi_0]$.

- Following spinor convention, the two components of a spin $1/2$ spinor $|\zeta\rangle = [\zeta_1, \zeta_0]$ have subscripts labelled by uppercase roman, A, B, etc.

6.2 Spinor geometry: flags, rotations and time-reversal

In this section we shall review various aspects of the geometry and algebra of spin $1/2$ spinors (hereafter called elementary or atomic spinors), that are needed in later sections. The details are easily found in mathematics textbooks (for example [Wig59, Cor53, LL77, PR84a, Nee97]), but may not be particularly familiar to readers here (sufficiently unfamiliar to be relegated to an appendix). No reference to spinor fields will be made in this section. The elementary spinors are all taken to be normalised, $\langle\zeta|\zeta\rangle = 1$.

At the heart of spinor geometry is the relation between complex two-dimensional vectors and real three-dimensional vectors, or, more accurately, between the Lie groups $SU(2)$ and $SO(3)$. The natural group map between the two Lie groups $SU(2) \longrightarrow SO(3)$ is $2 \to 1$, with $\pm u \mapsto o$ for $u \in SU(2)$, $o \in SO(3)$; this ambiguity of sign is important when considering half integer (fermionic) spins.

In section A.3, an elementary two-state spinor is related to a point in the complex plane including ∞ by (A.3.7), and so can be written as a unit vector $\mathbf{u}(\zeta)$ by stereographic projection (A.4.2). Therefore, with spherical polar angles θ, ϕ,

$$|\zeta\rangle = |\zeta(\theta, \phi)\rangle = \begin{bmatrix} \zeta_1 \\ \zeta_0 \end{bmatrix} = \begin{bmatrix} \cos\theta/2 \exp(-\mathrm{i}\phi/2) \\ \sin\theta/2 \exp(\mathrm{i}\phi/2) \end{bmatrix}, \tag{6.2.1}$$

and in the complex plane

$$z = z(\zeta) = \frac{\zeta_0}{\zeta_1} = \tan\theta/2 \exp(\mathrm{i}\phi). \tag{6.2.2}$$

Thus, up to a phase, the spinor $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ points in the $+x_3$ direction (at the north pole, *spin up*), corresponding to the origin of the complex plane and is denoted $|\uparrow\rangle$. The spinor $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ points towards the south pole (*spin down*) and is denoted $|\downarrow\rangle$. It stereographically projects to ∞. Other authors (such as [PR84a]) project from the north, rather than the south pole, but we find the sense in which orientation is preserved preferable.

The arguments of the spinor in (6.2.1) are such that $\arg(\zeta_0\zeta_1) = 0$, and is clearly normalised:

$$\langle\zeta|\zeta\rangle = \zeta_A^*\zeta_A = |\zeta_0|^2 + |\zeta_1|^2 = 1. \tag{6.2.3}$$

The normalisation and phase condition imply that for the atomic spinor (6.2.1), stereographic projection is a 1-1 correspondence. In quantum mechanics, the direction (θ, ϕ) of the spinor gives the direction (in space) of the quantum spin.

A natural question to ask is whether multiplying the spinor by a phase $\exp(-\mathrm{i}\chi/2)$ has any geometric significance. In fact it does; the spinor is said to represent a *flag* [PR84a], a unit vector (spherical angles (θ, ϕ) the 'flagpole') and a direction ('flag direction') χ perpendicular to the flagpole. The flagpole is given by the unit vector $\mathbf{u}(\zeta)$. The spinor in (6.2.1) is therefore modified:

$$|\zeta\rangle = |\zeta(\chi, \theta, \phi)\rangle = \begin{bmatrix} \cos\theta/2 \exp(-\mathrm{i}(\phi+\chi)/2) \\ \sin\theta/2 \exp(\mathrm{i}(\phi-\chi)/2) \end{bmatrix}. \tag{6.2.4}$$

where the *flag phase* χ of the spinor is defined in general by $-\arg\zeta_0\zeta_1$. The normalised flag spinor has three angle parameters χ, θ, ϕ. It is the first column of the unitary matrix

$$R = R(\chi, \theta, \phi) = \begin{bmatrix} \cos\theta/2 \exp(-\mathrm{i}(\phi+\chi)/2) & -\sin\theta/2 \exp(-\mathrm{i}(\phi-\chi)/2) \\ \sin\theta/2 \exp(\mathrm{i}(\phi-\chi)/2) & \cos\theta/2 \exp(\mathrm{i}(\phi+\chi)/2) \end{bmatrix}, \tag{6.2.5}$$

which represents, under the homomorphism SU(2)—→SO(3), rotation by the three Euler angles (rotate by χ about x_3, then by θ about $-x_2$, then by ϕ about x_3; see, for example, [Alt86]). Any element of SU(2) may be written in the form (6.2.5), just as any rotation may be written in terms of Euler angles (the three parameters being angles is what makes this Lie group compact). Another way of writing a general unitary transformation in spinor space, equivalent to a three-dimensional rotation, is the rotation by angle τ about an axis with (unit) direction $\mathbf{n}$,

$$R(\tau, \mathbf{n}) = \mathbf{1}\cos\tau/2 - \mathrm{i}\boldsymbol{\sigma}\cdot\mathbf{n}\sin\tau/2 = \exp(-\mathrm{i}\tau\boldsymbol{\sigma}\cdot\mathbf{n}/2), \tag{6.2.6}$$

where $\mathbf{1}$ is the 2×2 identity matrix and $\boldsymbol{\sigma}$ is the vector of Pauli matrices defined in (A.4.4) [Alt86]. Although this second representation may be more aesthetically pleasing [Fra88], we shall usually use the Euler angle representation. The flag spinor (6.2.4) therefore represents a rotation, and $|\zeta(\chi,\theta,\phi)\rangle = R(\chi,\theta,\phi)|\uparrow\rangle$. Since R represents a rigid rotation, the spinor $|\tilde{\zeta}\rangle$ representing the antipodal point on the sphere to $|\zeta\rangle$ is easy to find, being the result of the operation of R on $|\downarrow\rangle$:

$$|\tilde{\zeta}\rangle = |\tilde{\zeta}(\chi,\theta,\phi)\rangle = R(\chi,\theta,\phi)|\downarrow\rangle = \begin{bmatrix} -\sin\theta/2\exp(-\mathrm{i}(\phi-\chi)/2) \\ \cos\theta/2\exp(\mathrm{i}(\phi+\chi)/2) \end{bmatrix}. \tag{6.2.7}$$

The fact that $|\tilde{\zeta}\rangle$ is antipodal to $|\zeta\rangle$ is easily verified from the substitution

$$|\tilde{\zeta}\rangle = |\zeta(-\chi\pm\pi, \pi-\theta, \phi\pm\pi)\rangle. \tag{6.2.8}$$

The $\pm\pi$ terms here correspond to an overall sign ambiguity in $|\tilde{\zeta}\rangle$. We shall treat it as fixed by (6.2.7). The antipodal state therefore has the sense of χ reversed. The spinor $|\tilde{\zeta}\rangle$ shall be called the *dual* of the spinor $|\zeta\rangle$.

Rewriting R in terms of its Cayley-Klein parameters α, β [Alt86], its action on an arbitrary spinor $|\zeta\rangle$ gives another spinor:

$$R|\zeta\rangle = \begin{bmatrix} \alpha^* & -\beta^* \\ \beta & \alpha \end{bmatrix} \begin{bmatrix} \zeta_1 \\ \zeta_0 \end{bmatrix} = \begin{bmatrix} \alpha^*\zeta_1 - \beta^*\zeta_0 \\ \beta\zeta_1 + \alpha\zeta_0 \end{bmatrix}, \tag{6.2.9}$$

inducing a special unitary Möbius transformation on $z = \zeta_0/\zeta_1$ [Nee97],

$$z \to \frac{\alpha\zeta_0 + \beta\zeta_1}{\alpha^*\zeta_1 - \beta^*\zeta_0} = \frac{\alpha z + \beta}{\alpha^* - \beta^* z}. \tag{6.2.10}$$

The effect on z of taking the antipodal map $z(\zeta) \to z(\tilde{\zeta}) \equiv \tilde{z}$ is to conjugate, multiply by -1 and reciprocate (ie $\tilde{z} = -1/z^*$).

The fundamental bilinear form on spinors is the antisymmetric product [PR84a, Cor53], written

$$\varepsilon_{AB}\zeta_A\zeta'_B = \zeta_0\zeta'_1 - \zeta_1\zeta'_0, \tag{6.2.11}$$

with ε_{AB} the antisymmetric symbol $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$. This form is easily seen to be invariant with respect to rotations by R,

$$
\begin{aligned}
\varepsilon_{AB}R_{AC}\zeta_C R_{BD}\zeta'_D &= (\alpha\zeta_0 + \beta\zeta_1)(-\beta^*\zeta'_0 + \alpha^*\zeta'_1) - (-\beta^*\zeta_0 + \alpha^*\zeta_1)(\alpha\zeta'_0 + \beta\zeta'_1) \\
&= (|\alpha|^2 + |\beta|^2)\zeta_0\zeta'_1 - (|\alpha|^2 + |\beta|^2)\zeta_1\zeta'_0 \\
&= \varepsilon_{AB}\zeta_A\zeta'_B.
\end{aligned}
\tag{6.2.12}
$$

This invariance means that the antisymmetric product (hereafter called the *spinor product*, to distinguish from the inner product) has a geometric meaning, explained by [PR84a] pages 59-61 - its modulus is simply the sine of half the angle between the two flagpoles, the phase is half the sum of the two flag phases with respect to the geodesic line on the sphere joining the two flags, as shown in figure (6.1). The flag angles are measured with respect to the orientation imposed on the geodesic by the direction of the spinor product (from ζ to ζ' in (6.2.11)). If the order in the product were reversed, then π would be added to (or subtracted from) each flag angle, changing the phase by π, as expected from the antisymmetry of the spinor product. Obviously the spinor product of a spinor with itself is zero, and the product with its dual is 1. If each spinor in the product (6.2.11) is multiplied by a phase $\exp(i\gamma)$, the phase of the spinor product changes by $\exp(2i\gamma)$. The spinor product is crucial for defining C lines in the spinor formulation.

The usual hermitian inner product (the Hilbert space inner product, where the first term is conjugated) is clearly invariant with respect to rotations, by unitarity of R :

$$\langle\zeta|R^\dagger R|\zeta'\rangle = \langle\zeta|\zeta'\rangle = \zeta_0^*\zeta'_0 + \zeta_1^*\zeta_1. \tag{6.2.13}$$

The modulus of the inner product is the cosine of half the angle between the two flagpoles, and the phase is half of the difference between the angles the flags make with the geodesic. If the order of the spinors in the product is reversed, the phase becomes the negative of its previous value (the inner product is hermitian, so changing the order is equivalent to conjugating). The inner product of the spinor with itself is 1, and is zero with its dual,

$$\langle\zeta|\tilde{\zeta}\rangle = 0. \tag{6.2.14}$$
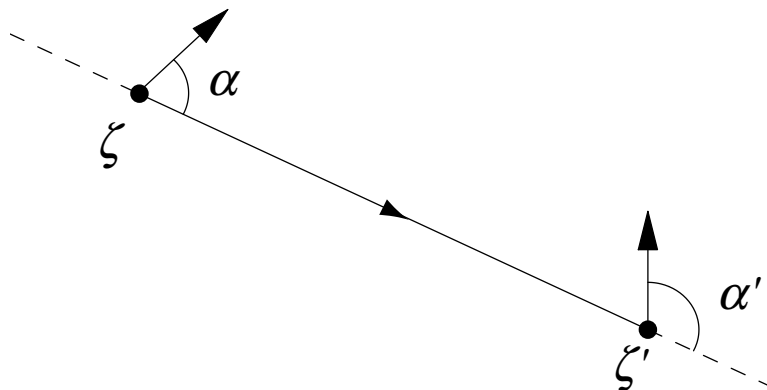
Figure 6.1: Calculating the phase of the spinor product $\varepsilon_{AB}\zeta_A\zeta'_B$. The directions of the flags are α, α' with respect to to the geodesic line joining the two spinors (represented here as a straight line), directed by the order of the product.

Using the inner product, it is possible to write the spinor product in bra-ket notation, as the matrix element of the time reversal operator $\hat{T}$ [Wig59, Sak94], defined

$$\hat{T} \equiv \hat{C}^\dagger \mathrm{i}\sigma_2, \qquad (6.2.15)$$

where σ_2 is the second Pauli spin matrix (A.4.4), $\hat{C}$ is the (right-acting) charge conjugation operator and operator notation $\hat{\bullet}$ is used for $\hat{C}, \hat{T}$ since conjugation cannot be written as a matrix. The adjoint indicates that conjugation acts to the left. The bra is therefore doubly conjugated, and $\mathrm{i}\sigma_2$ is simply the antisymmetric ε_{AB}, ie

$$
\begin{aligned}
\langle\zeta|\hat{T}|\zeta'\rangle &= \langle\zeta|\hat{C}^\dagger \mathrm{i}\sigma_2|\zeta'\rangle \\
&= \langle\zeta^*| \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} |\zeta'\rangle \\
&= \zeta_0\zeta'_1 - \zeta_1\zeta'_0.
\end{aligned}
\qquad (6.2.16)
$$

The time reversal operator is antiunitary ($\hat{T}\hat{T}^\dagger = -1$), and enables the operation of taking the dual to be put into operator form

$$
\begin{aligned}
|\tilde{\zeta}\rangle &= \hat{C}^\dagger\hat{T}|\zeta\rangle \\
&= \mathrm{i}\sigma_2|\zeta\rangle.
\end{aligned}
\qquad (6.2.17)
$$

This shows that, for instance, the spinor product between two spinors is minus that between their duals. The definition of $\hat{T}$ comes from the CPT theorem of quantum field

theory [SW64], which states that the operations of charge conjugation $\hat{C}$, parity reversal $i\sigma_2$ and time reversal leave a state invariant.

6.3 The Majorana representation of spin and polarization

Mathematically, the spinor algebra of the last section (in particular the use of unitary rotation matrices) is a realisation of the two-dimensional irreducible representation of the rotation group SO(3) generated by the Pauli matrices [Car66]. In fact, there is an irreducible representation for every dimension, familiar from quantum mechanics as the appropriate algebra for spin s, where the dimension of the representation is $n + 1 = 2s + 1$ (from now on, $n \equiv 2s$). The conventional choice for the spin matrices $\mathbf{S}^{(n)} = (S_1^{(n)}, S_2^{(n)}, S_3^{(n)})$ is for $S_3^{(n)}$ to be diagonal, and $S_1^{(n)}, S_2^{(n)}$ constructed from linear combinations of ladder operators acting on $S_3^{(n)}$ [VMK88].

The $n + 1$-dimensional complex vector space on which these matrices act is called *spin space* (*spin-s space*). Generalising the basis $|\uparrow\rangle, |\downarrow\rangle$ for atomic spinors, spin-s space has a basis

$$|\underbrace{\uparrow\uparrow\cdots\uparrow}_{n}\rangle, |\underbrace{\uparrow\uparrow\cdots\uparrow}_{n-1}\downarrow\rangle, \ldots, |\uparrow\underbrace{\downarrow\downarrow\cdots\downarrow}_{n-1}\rangle, |\underbrace{\downarrow\downarrow\cdots\downarrow}_{n}\rangle, \tag{6.3.1}$$

orthonormal with respect to the usual inner product. The basis is symmetrised with respect to all possible orderings of $\uparrow, \downarrow$; as stated in 6.1, notation is simplified by relabelling the basis spinors

$$|\mu\rangle \equiv |\underbrace{\uparrow\cdots\uparrow}_{\mu}\underbrace{\downarrow\cdots\downarrow}_{n-\mu}\rangle. \tag{6.3.2}$$

A general spinor of spin s, $|\psi\rangle$ (assumed normalised), is written in components (with summation convention, with μ labels running from 0 to n)

$$|\psi\rangle = \psi_\mu |\mu\rangle. \tag{6.3.3}$$

The $\uparrow, \downarrow$ atomic spinors in $|\mu\rangle$ (6.3.2) are formally indistinguishable, and spin-s space is mathematically the symmetric product of n copies of spin-1/2 space. It can be shown [FLS63b, Sak94] that a rotation of a $|\mu\rangle$ state is effectively a rotation of the $n + 1$ atomic spinors $\{|\uparrow\rangle, |\downarrow\rangle\}$ of which $|\mu\rangle$ is comprised, giving rise to the spin s rotation operator (in obvious generalisation of (6.2.6))

$$R^{(n)}(\tau, \mathbf{n}) = \exp(-i\tau\mathbf{S}^{(n)} \cdot \mathbf{n}). \tag{6.3.4}$$

The procedure for finding this matrix using Euler angles is very involved; the χ (and ϕ) matrices are easily found from (6.3.4) to be $\exp(-\mathrm{i}\chi S_3^{(n)})$ but the θ one is tricky, and its construction is described very clearly in [FLS63b] chapter 18, using (implicitly) the Clebsch-Gordan coefficients of group theory.

The time reversal operator similarly generalises to high spin [Wig59, Sak94],

$$\hat{T}^{(n)} = \hat{C}^{\dagger}\mathrm{i}\exp(-\mathrm{i}\pi S_2^{(n)}), \tag{6.3.5}$$

which may also be written in terms of spin $1/2$ σ_2 operations on the elementary spinor components of $|\mu\rangle$. Therefore

$$\hat{C}^{\dagger}\hat{T}|\mu\rangle = \mathrm{i}^{2\mu-n}|n-\mu\rangle \tag{6.3.6}$$

Up to a complex constant, all $\uparrow$ in $|\mu\rangle$ become $\downarrow$, and vice versa. The generalised spinor product is therefore (removing an overall factor i^{-n})

$$\langle\psi|\hat{T}|\psi\rangle = \sum_{\mu=0}^{n}(-1)^{\mu}\psi_{n-\mu}\psi_{\mu}, \tag{6.3.7}$$

which is invariant under rotations (6.3.4).

A geometric representation of an arbitrary $|\psi\rangle$ state is provided by the *Majorana representation* [Maj32, Pen94], which is now described. The basis of the spin space $\{|0\rangle, \ldots |n\rangle\}$ may be rewritten as a basis of monomials in an indeterminate $-z$,

$$|\mu\rangle \rightarrow (-1)^{\mu}\binom{n}{\mu}^{1/2}z^{\mu}. \tag{6.3.8}$$

The binomial factor $\binom{n}{\mu}^{1/2}$ is present because $|\mu\rangle$ is a symmetric product of $\mu \uparrow$ and $n-\mu$ $\downarrow$ atomic spinors, and is required for normalisation. Therefore, associated with any $|\psi\rangle$ is a *Majorana polynomial $p(\psi)$*,

$$\begin{aligned} p(\psi) &= (-1)^n\psi_n z^n + (-1)^{n-1}\sqrt{n}\psi_{n-1}z^{n-1} + \cdots + \psi_0 \\ &= \sum_{\mu=0}^{n}(-1)^{\mu}\binom{n}{\mu}^{1/2}\psi_{\mu}z^{\mu}. \end{aligned} \tag{6.3.9}$$

If z is regarded as a complex variable, this polynomial has n complex roots by the fundamental theorem of algebra, and is factorised,

$$p(\psi) = \psi_n\prod_{\overline{\mu}=1}^{n}(z_{\overline{\mu}} - z), \tag{6.3.10}$$

where the index label $\overline{\mu}$ is used rather than μ when the index runs from 1 to n. The complex values of the roots, as solutions to the equation

$$p(\psi) = 0, \tag{6.3.11}$$

are unaffected by the multiplication of $|\psi\rangle$ by any nonzero complex number, and so are invariant up to overall amplitude and phase. This was also the case with the polarization ellipse representation for complex vectors in chapter 4. All of the spin s polarization information is contained within the set of n complex roots $\{z_1, \ldots, z_n\}$, which are the coordinates of a point in complex projective space [Ati01].

Geometrically, each root $z_{\overline{\mu}}$ can be stereographically projected to a unit vector $\mathbf{u}_{\overline{\mu}}$ with angles $\theta_{\overline{\mu}}, \phi_{\overline{\mu}}$ associated with a (flagless) atomic spinor $|\zeta_{\overline{\mu}}\rangle$,

$$z_{\overline{\mu}} \rightarrow |\zeta_{\overline{\mu}}\rangle = \begin{bmatrix} \cos\theta_{\overline{\mu}}/2 \, \exp(-\mathrm{i}\phi_{\overline{\mu}}/2) \\ \sin\theta_{\overline{\mu}}/2 \, \exp(\mathrm{i}\phi_{\overline{\mu}}/2) \end{bmatrix}. \tag{6.3.12}$$

Any spinor $|\psi\rangle$ can therefore be decomposed into n atomic spinors $|\zeta_{\overline{\mu}}\rangle$, unlabelled since the roots $z_{\overline{\mu}}$ of the polynomial cannot be labelled (and may be permuted about a closed loop in parameter space). Geometrically, the spinor is described by the n root vectors $\mathbf{u}_{\overline{\mu}}$; these are usually called *principal normal directions* in spinor theory [Syn58, PR84a, PR84b]. It is readily verified that the configuration of n vectors rigidly rotates under the corresponding spatial rotation matrix when the spinor $|\psi\rangle$ is rotated, and so provides a physical picture of the spinor independent of the choice of coordinates. The mathematical proof of these claims is given in [PR84a] page 162. This representation of spin is called the *Majorana sphere* (because the unit vectors can be thought of as $n = 2s$ points on a unit sphere), and for spin s the Majorana sphere is denoted $\mathcal{M}_n$.

The Majorana sphere is a generalisation of the Riemann sphere for an atomic spinor $|\zeta\rangle$, whose Majorana polynomial is linear,

$$\zeta_0 - \zeta_1 z = 0. \tag{6.3.13}$$

The solution of (6.3.13) is (6.2.2), and it was for this reason that $-z$ was used rather than $+z$ in the monomials (6.3.8). ([Pen94] used the opposite convention.)

The Majorana polynomial is always considered to be of order n; if $\psi_n = 0$, it is understood that there is a root at ∞ (with root spinor $|\downarrow\rangle$), with possibly repeated root there if $\psi_{n-1} = 0, \psi_{n-2} = 0$, etc as well. In particular, this is the case for the Majorana

sphere representation of each basis spinor $|\mu\rangle$, related up to a constant to the monomial z^μ. This has μ roots at 0 and $n - \mu$ roots at ∞, so on the Majorana sphere has μ root spinors $|\uparrow\rangle$, and $n - \mu$ root spinors $|\downarrow\rangle$. The behaviour of the root vectors is quite complicated when two spin states are added, since the roots of the sum of Majorana polynomials is nonlinearly related to the roots of the summands.

Equation (6.3.13) shows that the Majorana sphere $\mathcal{M}_1$ is the same as the Riemann sphere introduced in A.3 for spin $1/2$; its connection to the Poincaré sphere and paraxial waves is described in section 6.4. $\mathcal{M}_2$ is equivalent to the nonparaxial polarization ellipse, as described by [Han98d], and is discussed in 6.7, and $\mathcal{M}_4$ is used to describe linear gravitational (and elastic) waves in section 6.8.

The Majorana polynomial (6.3.9) may be written in terms of the components of the root spinors $|\zeta_{\overline{\mu}}\rangle$,

$$p(\psi) = \kappa \prod_{\overline{\mu}=1}^{n} (\zeta_{\overline{\mu},0} - \zeta_{\overline{\mu},1} z), \tag{6.3.14}$$

where κ is a real constant ensuring that the state $|\psi\rangle$ associated with (6.3.14) remains normalised. The phase, $\frac{1}{2} \sum_{\overline{\mu}} \phi_{\overline{\mu}} + \arg \psi_n$, is distributed arbitrarily amongst the n root spinors as flags.

The dual polynomial $\tilde{p}(\psi) = p(\tilde{\psi})$, whose roots are antipodal to those of $p(\psi)$, is

$$\tilde{p}(\psi) = \sum_{\mu=0}^{n} \binom{n}{\mu}^{1/2} \psi_{n-\mu}^* z^\mu \tag{6.3.15}$$

(each coefficient ψ_μ in (6.3.9) is replaced by $(-1)^\mu \psi_{n-\mu}^*$) confirming that the operator $\hat{C}^\dagger \hat{T}$ in (6.3.6) dualises the spinor $|\psi\rangle$, ie

$$\tilde{p}(\psi) = p(\hat{C}^\dagger \hat{T} |\psi\rangle). \tag{6.3.16}$$

The generalised spinor product (6.3.7) therefore is once again the inner product of a state with its dual.

This section is concluded with reference to a result of Hannay [Han98b], which gives the expectation value $\langle \mathbf{S} \rangle$ of the spin operator geometrically in terms of the root vectors $\mathbf{u}_{\overline{\mu}}$. It is a complicated formula ([Han98b] equation (21) has a small typographical error), involving the sum over permutations of the root vectors. In the case of spin 1, with root vectors $\mathbf{u}_1, \mathbf{u}_2$, it reduces to the simple form [Han98d]

$$\langle \mathbf{S} \rangle = \frac{2(\mathbf{u}_1 + \mathbf{u}_2)}{3 + \mathbf{u}_1 \cdot \mathbf{u}_2} \tag{6.3.17}$$

which will be compared with $\mathbf{S}$ defined in (4.4.10) in section 6.7.

6.4   Plane waves and paraxial spin fields

It is a general result from relativistic field theory and the representation theory of the inhomogeneous Lorentz group (Poincaré group), that for a plane wave with wavevector $\mathbf{k}$ corresponding to a massless particle, choosing $\mathbf{k} = (0, 0, k)$, only components of the basis states $|0\rangle, |n\rangle$ can be nonzero. The result was first proved by Wigner [Wig39]. For spins 1 and 2, this condition is equivalent to transverseness; in the language of section 6.10, with cartesian coordinates $i, j = 1, 2, 3$,

$$\partial_i E_i = 0,$$
$$\partial_i h_{ij} = 0, \qquad (6.4.1)$$

This is because the only space dependence of a plane wave is the factor $\exp(\mathrm{i}k_i x_i)$, and so

$$k_i E_i = 0,$$
$$k_i h_{ij} = 0. \qquad (6.4.2)$$

This is only satisfied by the states $|0\rangle, |n\rangle$, which follows explicitly for $s = 1, 2$ from the forms given in 6.10. The conditions (6.4.1) hold for general superpositions of plane waves (such a superposition of vector waves was used in chapter 4). All wavefields we consider shall have this property. Such transverse fields correspond to massless particles [Wig39, BW48]; all of the fields' Fourier components are superpositions of helicity eigenstates.

In this section we shall examine the generalisation to higher spin of the paraxial vector wavefields of section 4.2. We begin by examining transverse plane waves more closely.

In paraxial fields, the propagation direction is fixed, and, by transverseness, there are only two basis states the spin vector may be in; $|\psi\rangle$ is effectively two-dimensional (in the paraxial approximation):

$$|\psi\rangle = \psi_0|0\rangle + \psi_n|n\rangle = \begin{bmatrix} \psi_n \\ \psi_0 \end{bmatrix}. \qquad (6.4.3)$$

In this way a spinor plane wave with well-defined propagation direction behaves like it has spin $1/2$; this is exactly the relation between the Poincaré sphere of A.4 and the Riemann sphere of A.3.

A plane wave is *right circularly polarized* if ψ_0 in (6.4.3) is 0, and left circularly polarized if $\psi_n = 0$. The wave is *linearly polarized* if $|\psi_0| = |\psi_n| = 1/\sqrt{2}$, with azimuthal angle of polarization dependent on the difference in arguments. The rotation operator about the

axis defined by $\mathbf{k}$ by an angle τ, on the two-dimensional helicity space defined by (6.4.3) is

$$R(\tau, \mathbf{k}) = \begin{bmatrix} \exp(-\mathrm{i}s\tau) & 0 \\ 0 & \exp(\mathrm{i}s\tau) \end{bmatrix}, \tag{6.4.4}$$

since the 3-spin operator $S_3^{(n)} = \mathrm{diag}[s, \ldots, -s]$. Any rotation around the propagation direction by $2\pi/s$ leaves a plane wave of spin s invariant, and states of linear polarization are orthogonal if related by a rotation by $\pi/2s$. For $s = 1$ this gives the familiar vector behaviour.

The relation between the Riemann and Poincaré spheres, and more generally Riemann and Majorana spheres, comes from writing the spinor $\begin{bmatrix} \psi_n \\ \psi_0 \end{bmatrix}$ on the Poincaré sphere with spherical angles θ_P, ϕ_P (cf (A.4.2))

$$\frac{\psi_0}{\psi_n} = \tan \theta_P/2 \exp(\mathrm{i}\phi_P) \tag{6.4.5}$$

The Majorana description of this state is similar but not identical, with Majorana polynomial

$$p(\psi) = \psi_0 + (-1)^n \psi_n z^n, \tag{6.4.6}$$

with n roots

$$z_{\overline{\mu}} = -\left(\frac{\psi_0}{\psi_n}\right)^{1/n} \exp(\mathrm{i}\pi(1 + 2\mu)/n) \tag{6.4.7}$$

The roots all have the same modulus $|\psi_0/\psi_n|^{1/n}$, with arguments equally spaced. They therefore form a regular polygon with n vertices (n-gon), centred at the origin of the complex plane. On the Majorana sphere, the n root vectors lie equally spaced on a common line of latitude defined by the polar angle θ_M, which together with the azimuth angles $\phi_{M,\overline{\mu}} = \phi_M + 2\overline{\mu}\pi/n$ are related to the Poincaré sphere angles θ_P, ϕ_P,

$$\tan \theta_P/2 \;=\; \tan^n \theta_M/2$$
$$\phi_P \;=\; n\phi_M. \tag{6.4.8}$$

For circularly polarized states, all roots are at 0 or ∞, and all root vectors coincide at either the north or south pole (they are simply $|n\rangle, |0\rangle$). Polygons in the northern hemisphere correspond to right-handed polarization (with Stokes parameter from the Poincaré sphere $s_3 > 0$), and equivalently for the southern hemisphere. The polygons of linearly polarized states lie on the equator of the Majorana sphere, and an axial rotation by $2\pi/s$ just permutes the root vectors, not changing the state.

The above equations show that the Majorana representation of a state is the nth root of its Poincaré sphere representation (taking roots of the complex number z). The Poincaré sphere thus generalised to arbitrary spin shares the appealing properties of the Poincaré sphere (eg antipodal states are orthogonal), and agrees nicely with the Majorana representation (whose use becomes necessary for three dimensional fields). It also illustrates features of high spin states not present for spin 1 - a general three dimensional spin state (with roots at arbitrary positions) is not, in general, expressible as a plane wave, even given the freedom of choice of propagation direction. For instance, the state $|\uparrow\uparrow\downarrow\rangle$ has two roots at the north pole and one at the south pole; it is not an equilateral triangle for any normal direction. Spin 1 is an exception, since any two root vectors form a 2-gon with propagation direction in the direction of the vector sum, and three dimensional spin 1 fields are made up of ellipses, just as they are in the paraxial case. This does not hold for higher spins, the geometry of whose generic states is much more complicated for three dimensions than two.

Similarly, a linearly polarized state (represented by an equatorial n-gon, or more generally, root vectors equally spaced on a great circle) does not correspond, except for spin 1, to any spin basis vector $|\mu\rangle$ for any axis; a 2-gon with vectors antipodal corresponds to the state $|1\rangle$ about the direction in which the vectors lie. The result is that, although paraxial spin s singularities (yet to be defined) correspond closely to their spin 1 analogues, the relationship between two and three dimensional spin singularities is rather different.

Paraxial spin s fields are very much like the paraxial vector fields of section 4.2. The paraxial field is defined, in comparison with (6.4.3), where $\mathbf{x} = (x_1, x_2)$,

$$|\psi_{\mathrm{P}}(\mathbf{x})\rangle = \begin{bmatrix} \psi_{\mathrm{P}n}(\mathbf{x}) \\ \psi_{\mathrm{P}0}(\mathbf{x}) \end{bmatrix} = \psi_{\mathrm{P}0}(\mathbf{x})|0\rangle + \psi_{\mathrm{P}n}(\mathbf{x})|n\rangle, \qquad (6.4.9)$$

mathematically identical to (4.2.1) in a circular basis. The components $\psi_{\mathrm{P}0}$, $\psi_{\mathrm{P}n}$ individually satisfy some wave equation, as in chapter 4.

There are therefore left and right handed C points, L lines and disclination points (where $\mathrm{Re}[\exp(-i\chi)(\psi_{P,0} - \psi_{P,n}, -i(\psi_{P,0} + \psi_{P,n}))]$ vanishes for phase χ) in paraxial high spin fields. The C points, from the n-fold symmetry of plane waves, are of index $\pm 1/n = \pm 1/2s$: it is likely that the lemon, star and monstar have some natural geometric interpretation for general spin, and most other properties generalise directly. The paraxial random wave model of 4.5 used the Poincaré sphere in constructing a random superposition of plane waves, as easily in a circular as cartesian basis. Obviously, the statistical

distribution of polarizations on the Poincaré sphere is independent of the spin of the wave, so the results of the statistical densities of C points, L lines and disclinations of 4.5 apply to paraxial waves of any spin.

The case for nonparaxial spin fields is entirely different, and the remainder of this chapter is devoted to a description of their nature.

6.5   C lines in three dimensional spin fields

We now generalise the codimension 2 C and L singularities of 4.3 to arbitrary fields of spin $s > 1/2$, using the Majorana description of spin states. The fields considered initially are spinor functions of space and possibly time $|\psi(\mathbf{x}, t)\rangle$, the components of which satisfy some wave equation. Such fields can be made from superpositions of infinitely many transverse plane waves, with arbitrary (random) directions, polarizations and phases. As well as identifying the singularities, we shall find natural functions of $|\psi(\mathbf{x})\rangle$ that vanish on the singularities, as found for the vector singularities of chapter 4.

C lines for spin 1 were simply places where the polarization was circular; that is, the two root vectors coincide. The codimensionality is expressible in terms of the spherical coordinates $(\theta_1, \phi_1), (\theta_2, \phi_2)$ of the two root vectors, as on a C line for spin 1 the following two conditions hold

$$\theta_1 = \theta_2, \qquad \phi_1 = \phi_2. \tag{6.5.1}$$

For an arbitrary spin s state, there are $n \equiv 2s$ unit vectors with coordinates $(\theta_{\overline{\mu}}, \phi_{\overline{\mu}})$, $\overline{\mu} = 1, \ldots, n$ on the Majorana sphere. Clearly, if $s > 1$, circular polarization requires all of these to be equal. This has more conditions than just the two we want for line structures in space (there are $2(n-1)$). However, if (6.5.1) is naively generalised so that only two of the n root vectors coincide,

$$\theta_{\overline{\mu}} = \theta_{\overline{\nu}}, \qquad \phi_{\overline{\mu}} = \phi_{\overline{\nu}}, \qquad \overline{\mu} \neq \overline{\nu}, \ \overline{\mu}, \overline{\nu} \in \{1, \ldots, n\}, \tag{6.5.2}$$

there are only two requirements, so (6.5.2) ought to be satisfied along lines. Although there are $n(n-1)/2$ different possibilities of different roots coalescing, since the roots cannot be globally labelled, all such different possibilities are identified.

The two conditions of (6.5.2) show that such generalised C lines (C for Coalescent roots or Coincident root vectors) occur with codimension two, and are lines in space (although, of course, the direction in which the root vectors coincide is independent of the direction

of the C line). In a small loop in space around the C line the root vectors which coincide are permuted. It is easy to verify that this is the case with spin 1; the polarization ellipse around the singularity rotates by π, being index $1/2$. This implies that generalised C lines have index $1/2$, rather than index $1/n$ (which is the case for generalised paraxial C points).

The index $1/2$ nature may be understood further by considering what happens to the two coinciding vectors in a transverse neighbourhood to the C line; their positions may be expressed by the line element connecting them, which is necessarily undirected since the root vectors cannot be labelled. The line disappears at the singularity, and the line element rotates by $\pm\pi$ in a small loop around the singularity, as is usual with line fields, and may have a lemon, star or monstar pattern.

A state on the Majorana sphere is mathematically defined as the symmetric product of n copies of the sphere $\mathcal{S}_2$, which is equivalent to the direct product of n spheres factored by the group of permutations on n objects Σ_n, ie

$$\prod_n {}_{\mathrm{sym}}\mathcal{S}_2 \simeq \left(\prod_n \mathcal{S}_2\right)/\Sigma_n, \tag{6.5.3}$$

reflecting the fact that the root vectors cannot be labelled. Traversing a closed loop in $\mathbf{x}$ space, the n vectors may be permuted. Now, we have seen that two root vectors are permuted around a C line; clearly if the circuit does not thread any C lines, it may be smoothly (homotopically) contracted to a point without the roots coalescing, ie the roots may be labelled unambiguously, and no permutation has taken place. Therefore, generically, the C lines are the *generators* of the permutations; any permutation can be broken down into 2-cycles (two element permutations), although the 2-cycle decomposition of a more general permutation is not unique. The C lines are therefore the singularities of labelling the root vectors, and the roots in a singularity-free simply-connected volume of space may be unambiguously labelled.

In the language of topological defects in condensed matter (section 1.4), the permutation group acts like the fundamental group in order parameter space (that is, different loops in space correspond to different group elements). For any $n > 2$, the permutation group Σ_n is nonabelian, and the relevant abelian group of singularity indices is the group of cosets (which is abelian) [Mer79]; for permutation groups this is always the two element group Σ_2, so all C lines are topologically identified. The crossings of C lines are complicated, and the geometry depends whether three or four roots coincide at the crossing

point; more work needs to be done here to understand the morphology correctly.

One expects the significance of C lines to decrease as spins of higher s are considered, and the role of individual root vectors is less. However, due to the increasing number of real freedoms of such fields ($2n$ polarization freedoms), it is inevitable that a structure of only codimension 2 should be insensitive to many of the geometrical details.

It is desirable to express the complicated polarization structure using a complex scalar (homogeneous in the components ψ_μ), which vanishes on C lines, generalising the polarization scalar φ of (4.3.6). Such a scalar, if invariant with respect to rotations of the Majorana sphere, would be a crude (two real parameter) description of the polarization structure, but would detect, as phase singularities, the C singularity structure. In fact, such a scalar exists.

The *discriminant* of a polynomial $p = \sum_{\mu=0}^{n} a_\mu z^\mu$ is defined to be the product of squares of the differences of the roots,

$$(-1)^{n(n-1)/2} \prod_{1 \leq \overline{\mu} < \overline{\nu} \leq n} (z_{\overline{\mu}} - z_{\overline{\nu}})^2. \tag{6.5.4}$$

It is clearly invariant with respect to permutations of the root labels. It can be found in terms of the coefficients a_μ using a variety of techniques, including that of the bezoutian matrix $\mathcal{B}(p)$ of the polynomial [Meh89]. We modify the definition slightly in order to make the discriminant $\mathcal{D}$ homogeneous:

$$\mathcal{D}(p) = a_n^{2(n-1)}(-1)^{n(n-1)/2} \prod_{1 \leq \overline{\mu} < \overline{\nu} \leq n} (z_{\overline{\mu}} - z_{\overline{\nu}})^2. \tag{6.5.5}$$

The degrees of the summands in $\mathcal{D}$ are $2(n-1)$ (that is, a '$2n-2$'-ic form). Note that each root $z_{\overline{\mu}}$ appears $2(n-1)$ times in the product (6.5.5) (it appears twice with each of the other $n-1$ roots).

Since the discriminant is zero if and only if roots coincide, it generically vanishes exactly on C lines, and, in general, is a complex scalar, homogeneous in the field variables. It is a suitable polarization scalar, provided it can be shown to be invariant with respect to rotation of the Majorana sphere (equivalent to passive coordinate transformations).

The discriminant of a spin field $|\psi\rangle = \psi_\mu |\mu\rangle$ is the scalar field (using (6.3.10), (6.5.5)),

$$\varphi(\mathbf{x}) = \mathcal{D}(p(\psi(\mathbf{x}))) = \psi_n^{2(n-1)} \prod_{1 \leq \overline{\mu} < \overline{\nu} \leq n} (z_{\overline{\mu}} - z_{\overline{\nu}})^2. \tag{6.5.6}$$

Since each root $z_{\overline{\mu}}$ can be written in terms of its corresponding atomic spinor, $z_{\overline{\mu}} = \zeta_{\overline{\mu}0}/\zeta_{\overline{\mu}1}$, φ may be rewritten

$$\varphi = \frac{\psi_n^{2(n-1)}}{\prod_{\overline{\mu}} \zeta_{\overline{\mu}1}^{2(n-1)}} \prod_{\overline{\mu}<\overline{\nu}} (\zeta_{\overline{\mu}0}\zeta_{\overline{\nu}1} - \zeta_{\overline{\nu}0}\zeta_{\overline{\mu}1})^2. \tag{6.5.7}$$

The difference between each pair of roots is now the spinor product (6.2.11) between the root spinors $|\zeta_{\overline{\mu}}\rangle, |\zeta_{\overline{\nu}}\rangle$, and is invariant under rotations by (6.2.12), and the factor outside the product is equal to κ in (6.3.14). φ is a complex scalar invariant of $|\psi\rangle$, and is written more transparently

$$\varphi = \kappa^{2(n-1)} \prod_{\overline{\mu}<\overline{\nu}} \langle \zeta_{\overline{\mu}} | \hat{T} | \zeta_{\overline{\nu}} \rangle^2 \tag{6.5.8}$$

φ is homogeneous in the components of the spinor, and the factor of $\kappa^{2(n-1)}$ shows that, if $|\psi\rangle$ is multiplied by a phase $\exp(i\chi)$, φ changes by $2(n-1)$ times that phase $(\exp(2i\chi)$ in the case of $s=1$). We shall see in section 6.7 that for spin 1, the two definitions for φ are the same.

So far, no general geometric interpretation of φ for arbitrary spinors has been found. However, the nature of the singularity and its relation to φ can be described in terms of the root spinors. As with vectors, although the polarization geometry is invariant with respect to global phase (gauge) transformations, φ is dependent on the phase of the field; the phase can be globally transformed arbitrarily anywhere except where it is singular. The phase of the spinor can be considered as being distributed as flag phases amongst the n atomic spinors. If two flagpoles coincide, their spinor product is zero, and the phase, defined graphically in figure (6.1), cannot be defined. This is the case whatever the phase distribution amongst the atomic flags, so C lines are flag singularities of the root vectors.

6.6   L lines in three dimensional spin fields

With the same assumptions as the previous section, we now investigate how L lines in three dimensions are generalised for high spin. As with generalised C lines, there are too many restrictions (too high a codimension) for genuine linear polarization states to occur generically for $s > 1$ (fixing all of the root vectors to lie on an arbitrary equatorial n-gon would have codimension $2n - 3$).

We can look for relevant pairwise conditions for the root spinors to satisfy, as in the previous section. Since, for spin 1, L lines have the two root vectors antipodal on the

Majorana sphere, we impose this as the generalised L condition: an L singularity is a locus where any two of the Majorana root vectors are antipodal, or

$$\mathbf{u}_{\overline{\mu}} = -\mathbf{u}_{\overline{\nu}}, \qquad \overline{\mu} \neq \overline{\nu}, \ \mu, \nu \in \{1, \ldots, n\}. \tag{6.6.1}$$

Equivalently, using the polar angles of the root vectors,

$$\theta_{\overline{\mu}} = \pi - \theta_{\overline{\nu}}, \ \phi_{\overline{\mu}} = \phi_{\overline{\nu}} + \pi, \qquad \overline{\mu} \neq \overline{\nu}, \ \overline{\mu}, \overline{\nu} \in \{1, \ldots, n\}. \tag{6.6.2}$$

These two conditions show that L singularities (two root vectors coLlinear with origin) have codimension 2, as desired.

As with the spin 1 case, L lines have index 1, as can be seen from the local behaviour of the vector $\mathbf{v} = \mathbf{u}_{\overline{\mu}} + \mathbf{u}_{\overline{\nu}}$ where $\overline{\mu}, \overline{\nu}$ locally label the root vectors which are antipodal on the L line. $\mathbf{v}$ clearly vanishes there, and being a vector, its field zeros have integer index, generically ± 1. $\mathbf{v}$ is parallel to $\mathbf{N}$ of (4.3.4).

Since L lines have nothing to do with roots coinciding, L lines appear to be unrelated to permutation of the root vectors, and appear to be less interesting than C lines for general spin. As with C lines, L lines may cross, but in general the sets of antipodal root vectors are disjoint (unless there is also a C line crossing). As with C lines, the physical singularity is related to the fact that there is no unique geodesic between antipodal atomic spinors, and their inner product vanishes.

By analogy with (6.5.8), recalling that the inner product is invariant with respect to rotations (6.2.13) and zero for antipodal atomic spinors (6.2.14), we can define a real positive scalar invariant field

$$\lambda = \rho \prod_{\overline{\mu}, \overline{\nu}} \langle \zeta_{\overline{\mu}} | \zeta_{\overline{\nu}} \rangle. \tag{6.6.3}$$

(ρ acts like κ in (6.5.8)). This form is of degree $2n$, as it includes terms in the product where $\overline{\mu} = \overline{\nu}$. λ may be rewritten in terms of the Majorana polynomial roots,

$$\lambda = \prod_{\overline{\mu}, \overline{\nu}} (\zeta_{\overline{\mu}0}^* \zeta_{\overline{\nu}0} + \zeta_{\overline{\mu}1}^* \zeta_{\overline{\nu}1}) = \left(\prod_{\overline{\mu}, \overline{\nu}} \zeta_{\overline{\mu}0}^* \zeta_{\overline{\nu}1} \right) \prod_{\overline{\mu}, \overline{\nu}} \left(z_{\overline{\nu}} + \frac{1}{z_{\overline{\mu}}^*} \right) \tag{6.6.4}$$

which is the *resultant* $\mathcal{R}(p, \tilde{p})$ [Meh89] of the Majorana polynomial $p(\psi)$ and its dual $\tilde{p}(\psi)$, defined in (6.3.15).

Unlike φ in (6.5.8), λ in (6.6.4) is real and positive. In the spin 1 case, we shall see that it is equal to $\mathbf{N} \cdot \mathbf{N}$ with $\mathbf{N}$ as defined previously in (4.3.4), but a general geometric interpretation of λ has yet to be found. It is likely that it is some composite of a geometrically simpler object (like $\mathbf{N} \cdot \mathbf{N}$).

Before the discussion of arbitrary spin is concluded, we briefly consider the possibility that certain other geometrical configurations of the Majorana root vectors may give interesting singular structures not present for $s = 1$. An example might be three root vectors lying on an equilateral triangle, which would happen with codimension 2. Such configurations indeed would occur generically, but it is hard to see how their presence could be detected using scalars as simple as φ and λ, which are the natural bilinear forms on atomic spinors.

6.7   The Majorana interpretation for vector waves

Everything that has been said concerning generalised C and L lines, when $s = 1$, reduces to the formalism already discussed in section 4.3 for vector waves. Firstly, the configuration of the two Majorana root vectors $\mathbf{u}_1, \mathbf{u}_2$ has a simple interpretation in terms of the polarization ellipse (due to Hannay [Han98d]), which is demonstrated here using angular momentum.

The expectation of angular momentum $\mathbf{S}$, in units of $\hbar$, was stated in (6.3.17) to be $2(\mathbf{u}_1 + \mathbf{u}_2)/(3 + \mathbf{u}_1 \cdot \mathbf{u}_2)$. If the angle between the $\mathbf{u}_1$ and $\mathbf{u}_2$ is 2θ, (so each vector subtends an angle θ with the bisector), which is the angle θ_M for a plane wave defined above (6.4.8), then

$$|\mathbf{S}| = \frac{4\cos\theta}{3 + \cos 2\theta} = \frac{2\cos\theta}{1 + \cos^2\theta}. \tag{6.7.1}$$

Now, recalling (4.4.10), and using the notation of chapter 4,

$$|\mathbf{S}| = \frac{2|\mathbf{p} \wedge \mathbf{q}|}{p^2 + q^2} = \frac{2p_0 q_0}{p_0^2 + q_0^2}, \tag{6.7.2}$$

where $\mathbf{p_0}, \mathbf{q_0}$ are orthogonal by appropriate choice of phase. If p_0 is the major semiaxis, q_0 the minor, then (6.7.2) can be rewritten

$$|\mathbf{S}| = \frac{2q_0/p_0}{1 + q_0^2/p_0^2}, \tag{6.7.3}$$

then, equating (6.7.1), (6.7.3), the polarization ellipse has eccentricity ε,

$$\varepsilon = \sin\theta. \tag{6.7.4}$$

The geometry of this result is shown in figure (6.2).

In particular, (6.7.4) reproduces the fact that $|\mathbf{S}| = 1$ for circular polarization, and 0 for linear polarization. The Majorana sphere $\mathcal{M}_2$ therefore provides a parameterisation
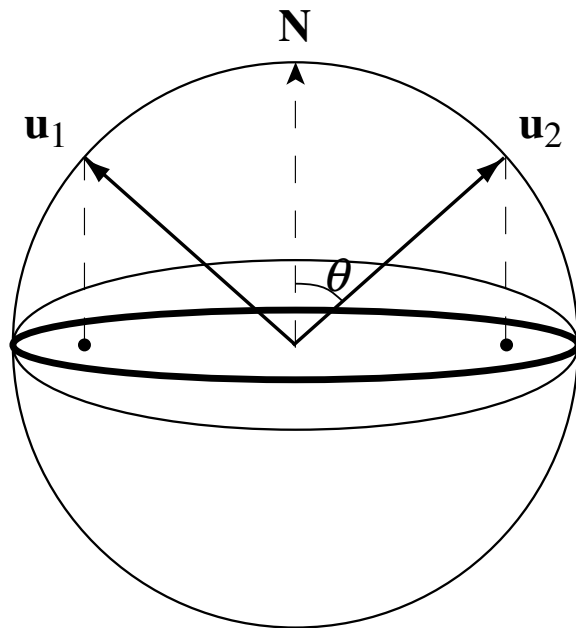
Figure 6.2: The geometric significance of equation (6.7.4): the projections of the two root vectors $\mathbf{u}_1, \mathbf{u}_2$ into the equatorial plane normal to their bisector are the foci for the polarization ellipse with major semiaxis 1.

for normalised, phaseless, oriented ellipses in space, just as the Poincaré sphere provides one for the plane.

The two definitions for φ (equations (4.3.6), (6.5.8)) are equivalent. The homogeneous discriminant of a quadratic equation is well known; with the Majorana binomial factor, the discriminant is

$$\varphi = \psi_1^2 - 2\psi_0\psi_2, \tag{6.7.5}$$

equivalent to the quadratic form with matrix $\left[\begin{smallmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{smallmatrix}\right]$. The unitary transformation between cartesian and spherical bases is constructed in 6.10. If the ψ components in (6.7.5) are transformed to components of $\mathbf{E} = (E_1, E_2, E_3)$, φ becomes

$$\begin{aligned} \varphi &= E_3^2 - 2\frac{1}{\sqrt{2}}(-E_1 - iE_2)\frac{1}{\sqrt{2}}(E_1 - iE_2) \\ &= E_1^2 + E_2^2 + E_3^2, \end{aligned} \tag{6.7.6}$$

agreeing with (4.3.6), corresponding to the quadratic form matrix $\left(\begin{smallmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{smallmatrix}\right)$.

Similarly, λ in (6.6.3) is proportional to $\mathbf{N} \cdot \mathbf{N}$, with $\mathbf{N} = \mathbf{p} \wedge \mathbf{q}$ (4.3.4).

The connection between complex vectors and spin 1/2 spinors was taken advantage in forms reminiscent to those here at the very beginning of spinor theory; Cartan [Car66]

notices the 1-1 relation between a spin 1/2 spinor and a null vector $\mathbf{E}$ ($\mathbf{E}\cdot\mathbf{E}=0$); the spin 1 spinor's coincident roots are mathematically equivalent to an atomic spinor. Synge [Syn58] (apparently independent of Majorana) also realised some of the geometry here in relation to the electromagnetic vector $\mathbf{V} = \mathbf{E} + \mathrm{i}\mathbf{H}$ (4.4.2), and coined the term *principal null directions* for the root vectors. As discussed in 4.4, the interpretation of the singularities for this field can be made in terms of Lorentz transformations, which are easily realised on the Majorana sphere via Möbius transformations.

6.8 Polarization singularities in gravitational waves

In this section we indicate how the material from the previous sections may be used to find the polarizations singularities in spin 2 waves, in particular (linearised) gravity waves. It also suggests the appropriate polarization structure of elastic tensor waves (the connection between the two having been investigated by Hayes [Hay84]).

The Majorana sphere $\mathcal{M}_2$ here is used to describe the polarization structure of complex traceless 3×3 symmetric matrices (the relation via spherical harmonics is outlined in section 6.10). As stated in the chapter introduction, it was realised too late that this relationship is well-known in the theory of solutions of the general Einstein field equations by classification of the Weyl tensor (equivalent to a traceless 3×3 symmetric complex tensor) into the Petrov types [Syn64] and its relation to principal normal directions by Penrose [Pen60, NP62, KSMH80, PR84b]. Assuming that most readers are unfamiliar with this material, the exposition shall be presented briefly, as originally envisaged; no references have been found in the literature to generic loci of codimension 2 where the Weyl tensor has a different Petrov type (four independent root vectors corresponds to type I in Petrov notation, two coalescent vectors to type II (or types 1, 2a in the notation of [Syn64])). However, the discussion shall not be particularly complete, nor shall L lines be investigated.

The usual starting point for studies of gravitational waves is the linearised Einstein equations (see, for instance, [Sch85, MTW73]); where μ, ν temporarily label spacetime indices, $T_{\mu\nu} = 0$ and

$$\nabla^2 g_{\mu\nu} = \partial_t^2 g_{\mu\nu}, \tag{6.8.1}$$

so each component satisfies the time-dependent wave equation (1.5.1). Gravitational disturbances to the otherwise flat metric are assumed small, and so differ only slightly from

the standard Minkowski metric $\eta_{\mu\nu}$:

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}. \tag{6.8.2}$$

$h_{\mu\nu}$ therefore satisfies (6.8.1), which is the standard relativistic wave equation for massless fields of spin 2 [Wig39, Bar47, BW48]. The appropriate choice of (gravitational) gauge is the so-called 'transverse-traceless' gauge [Sch85, MTW73], which chooses once and for all a global time direction (since the metric is almost flat), and $h_{\mu\nu}$ becomes h_{ij}, with space indices only, and satisfies

$$\partial_i h_{ij} = 0, \tag{6.8.3}$$

as discussed at equation (6.4.1). It also is traceless,

$$h_{ii} = 0. \tag{6.8.4}$$

We therefore have a rank 2 tensor analogue of the $\mathbf{E}$ field considered previously; h_{ij} can be decomposed into a Fourier sum of plane waves. The two orthogonal linear polarizations for a given propagation direction (we choose the 3-direction) are usually referred to as the '+' and '×' states, and have h_{ij} representations

$$\mathbf{e}_+ = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$\mathbf{e}_\times = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{6.8.5}$$

and homogeneous plane waves with these polarizations are

$$h_{\bullet ij} = e_{\bullet ij} \exp(\mathrm{i}(kx_3 - \omega_k t)), \tag{6.8.6}$$

where $\bullet = +, \times$. The effect of these two polarizations on a ring of test particles are shown in [Sch85] figure (9.1).

Right $(+)$ and left $(-)$ circularly polarized plane waves are therefore

$$\mathbf{e}_\pm = \mp(\mathbf{e}_+ \pm \mathrm{i}\mathbf{e}_\times) \exp(\mathrm{i}kx_3 - \omega_k t). \tag{6.8.7}$$

The inner product is defined as half the trace of the matrix product

$$\mathbf{h}^* \cdot \mathbf{h}' = \frac{1}{2} \operatorname{tr} \mathbf{h}^* \mathbf{h}' = \frac{1}{2} h_{ij}^* h_{ij}'. \tag{6.8.8}$$

An arbitrary sum of (complex) plane waves in arbitrary directions gives rise to a traceless symmetric matrix field with complex elements, written (where all space, time dependence is suppressed)

$$\mathbf{h} = \begin{pmatrix} U - V/\sqrt{3} & Z & Y \\ Z & -U - V/\sqrt{3} & X \\ Y & X & 2V/\sqrt{3} \end{pmatrix} \tag{6.8.9}$$

which is traceless and symmetric. This matrix is considered in 6.10, where it is expressed in terms of a 5-dimensional vector with complex components $\{U, V, X, Y, Z\}$. It is related to the components of the corresponding 2-spinor $|\psi(h)\rangle = [\psi_4, \psi_3, \psi_2, \psi_1, \psi_0]$, where

$$\mathbf{h} = \frac{1}{\sqrt{2}} \begin{pmatrix} \psi_4 - \sqrt{\frac{2}{3}}\psi_2 + \psi_0 & \mathrm{i}(\psi_4 - \psi_0) & \psi_1 - \psi_3 \\ \mathrm{i}(\psi_4 - \psi_0) & -\psi_4 - \sqrt{\frac{2}{3}}\psi_2 - \psi_0 & -\mathrm{i}(\psi_1 + \psi_3) \\ \psi_1 - \psi_3 & -\mathrm{i}(\psi_1 + \psi_3) & 2\sqrt{\frac{2}{3}}\psi_2 \end{pmatrix}. \tag{6.8.10}$$

The Majorana polynomial $p(h)$, by (6.3.9) is

$$p(h) = \psi_4 z^4 - 2\psi_3 z^3 + \sqrt{6}\psi_2 z^2 - 2\psi_1 z + \psi_0. \tag{6.8.11}$$

It has four roots, corresponding to the four root vectors of the Majorana sphere $\mathcal{M}_4$. The polarization scalar φ is the discriminant of this polynomial, a form of degree 6, and is given in terms of spin components $\psi_4, \ldots, \psi_0$ and cartesian components $U, \ldots Z$:

$$
\begin{aligned}
\varphi =\ & -3\psi_1^2\psi_2^2\psi_3^2 + 3\sqrt{6}\psi_0\psi_2^3\psi_3^2 + 8\psi_1^3\psi_3^3 - 9\sqrt{6}\psi_0\psi_1\psi_2\psi_3^3 + 27/2\,\psi_0^2\psi_3^4 + 3\sqrt{6}\psi_1^2\psi_2^3\psi_4 \\
& -18\psi_0\psi_2^4\psi_4 - 9\sqrt{6}\psi_1^3\psi_2\psi_3\psi_4 + 60\psi_0\psi_1\psi_2^2\psi_3\psi_4 + 3\psi_0\psi_1^2\psi_3^2\psi_4 - 8\psi_0^3\psi_4^3 \\
& -18\sqrt{6}\psi_0^2\psi_2\psi_3^2\psi_4 + 27/2\psi_1^4\psi_4^2 - 18\sqrt{6}\psi_0\psi_1^2\psi_2\psi_4^2 + 24\psi_0^2\psi_2^2\psi_4^2 + 24\psi_0^2\psi_1\psi_3\psi_4^2 \\
=\ & -U^6 + 6U^4V^2 - 9U^2V^4 - 3U^4X^2 + 9\sqrt{3}U^3VX^2 - 15U^2V^2X^2 - 3\sqrt{3}UV^3X^2 \\
& +15/2\,U^2\,X^4 - 9\sqrt{3}/2\,UVX^4 - 3/4\,V^2X^4 - X^6 - 3U^4Y^2 - 9\sqrt{3}U^3VY^2 \\
& -15U^2V^2Y^2 + 3\sqrt{3}UV^3Y^2 - 39/2\,U^2X^2Y^2 - 3/2\,V^2X^2Y^2 - 3X^4Y^2 \\
& +15/4\,U^2Y^4 + 9\sqrt{3}/2\,UVY^4 - 3/4\,V^2Y^4 - 3X^2Y^4 - Y^6 - 18\sqrt{3}U^2VXYZ \\
& +6\sqrt{3}V^3XYZ - 27UX^3YZ + 9\sqrt{3}VX^3YZ + 27UXY^3Z + 9\sqrt{3}VXY^3Z \\
& -3U^4Z^2 + 12U^2V^2Z^2 - 9V^4Z^2 - 6U^2X^2Z^2 + 9\sqrt{3}UVX^2Z^2 - 15V^2X^2Z^2 \\
& -3X^4Z^2 - 6U^2Y^2Z^2 - 9\sqrt{3}UVY^2Z^2 - 15V^2Y^2Z^2 + 21X^2Y^2Z^2 - 3Y^4Z^2 \\
& -18\sqrt{3}VXYZ^3 - 3U^2Z^4 + 6V^2Z^4 - 3X^2Z^4 - 3Y^2Z^4 - Z^6.
\end{aligned} \tag{6.8.12}
$$

What does this mean for the tensor (6.8.9), (6.8.10)? For convenience, we work in the spin basis, and write down the characteristic polynomial $q(t)$ of $\mathbf{h}$:

$$
\begin{aligned}
q(t) &= \det(h_{ij} - t\delta_{ij}) \\
&= -t^3 + 2t\psi_2^2 + \frac{4\sqrt{2}\,\psi_2^3}{3\sqrt{3}} - 4t\psi_1\psi_3 - 4\sqrt{\frac{2}{3}}\psi_1\psi_2\psi_3 + 4\psi_0\psi_3^2 \\
&\quad + 4t\psi_0\psi_4 + 4\psi_1^2\psi_4 - 8\sqrt{\frac{2}{3}}\psi_0\psi_2\psi_4 \\
&= -t^3 + t\left(U^2 + V^2 + X^2 + Y^2 + Z^2\right) - \frac{2U^2V}{\sqrt{3}} + \frac{2V^3}{3\sqrt{3}} - UX^2 + \frac{VX^2}{\sqrt{3}} \\
&\quad + UY^2 + \frac{VY^2}{\sqrt{3}} + 2XYZ - \frac{2VZ^2}{\sqrt{3}}.
\end{aligned}
\tag{6.8.13}
$$

The eigenvalues of h_{ij} (roots of (6.8.13)) sum to zero by tracelessness, and the discriminant of the cubic $q(t)$ is easily found to be *exactly the same* as φ, the discriminant of the quartic polynomial p, up to a factor; C lines are places where the matrix $\mathbf{h}$ is degenerate.

It is very surprising that these two polynomials, the quartic Majorana polynomial p, linear in the spin coefficients ψ_μ, and the cubic characteristic polynomial q, cubic in the ψ_μ, share the same discriminant. Even more astonishing, however, is that the relationship between the coefficients of p and q is exactly the same as those of Ferrari's construction of the solution to the quartic polynomial [Boy68] pages 286-287: in order to find the roots of a quartic polynomial, one constructs a cubic polynomial whose solution is found by the Cardan-Tartaglia formula, and then finds the roots of the quartic in terms of the roots of the cubic. The (hitherto) purely algebraic method for solving quartic equations therefore has a physical interpretation, where the two polynomials (6.8.11), (6.8.13) are related by a purely geometric construction. Maybe a reason for the this analogy is that the Ferrari method is indeed the only way a quartic and cubic polynomial may be related (with respect to roots), and the physics of C lines requires this connection to be made.

The codimension of degeneracies of complex symmetric matrices is indeed 2, as may be readily verified by considering the 2×2 symmetric matrix $\left(\begin{smallmatrix} \alpha & \gamma \\ \gamma & \beta \end{smallmatrix}\right)$, whose characteristic polynomial has discriminant

$$
(\alpha + \beta)^2 - 4(\alpha\beta - \gamma^2) = (\alpha - \beta)^2 + 4\gamma^2.
\tag{6.8.14}
$$

If α, β, γ are complex, both of the complex summands on the right hand side of (6.8.14) must vanish, as a normal phase singularity. If they are all real, degeneracies are still codimension 2 since to be zero requires the vanishing of two squares. By comparison, degeneracies of hermitian matrices have codimension 3 [Ber94c].

The precise meaning of degeneracies in the h wavefield, in terms of its real part as phase χ is varied,

$$h_{ij}^{(r)} = \operatorname{Re} h_{ij} \exp(-\mathrm{i}\chi) \tag{6.8.15}$$

is not clear. Each of the real matrices $h_{ij}^{(r)}$ is traceless, and if χ is changed by π, each of its eigenvalues $t_i \to -t_i$; so at any point there must be at leat two phases for which one of the eigenvalues vanishes. These structures are highly complicated, and have yet to be fully understood.

6.9   Discussion

Unfortunately, our explanation of tensor singularities finishes here; obviously much more work needs to be done before the understanding of these objects reaches the level of vector singularity understanding.

In this chapter, we have interpreted the vector polarization fields in terms of spin, found the significance of C and L lines in these spin fields, and generalised to spin fields with spin greater than 1, both for paraxial and nonparaxial fields. We have seen that the paraxial spin singularities are equivalent to the paraxial vector singularities of section 4.2, including their statistical densities in isotropic random paraxial spin fields, which are the same as those calculated in section 4.5. Questions as to the morphology of these singularities (generalised C points having index $1/n$) still remain, and no explicit examples of these have been studied in any particular detail.

The C and L singularities of spin fields in three dimensions are not particularly well understood. The topological nature of their interactions, the role of the permutation group, the possibility of other sorts of singularity (maybe with codimension other than 2) are all subjects requiring further investigation. This probably will not increase our understanding of spin 1 singularities, but may provide further tools in manipulating them.

We have only looked at C lines in spin 2 fields. These are topologically more interesting, and more fundamental; the discriminant is a simpler and more fundamental object than the resultant of a polynomial with its dual.

C and L lines in high spin fields may be subjected to statistical analysis, and the random wave model of section 4.6 generalises directly, using the plane waves described in section 6.4. The resulting random field has $U, \ldots, Z$ (equivalently $\psi_4, \ldots, \psi_0$) as five independent, identically distributed gaussians. However, the singularity density calculations

are intractable; expression (6.8.12), and its vorticity, cannot be simplified further.

Although the newly discovered literature on Petrov types must be digested and understood, it is likely that the similarities cease at this point. The key to understanding the singular nature of C lines in spin 2 fields is (6.8.15); as the phase varies the orthogonal eigenframe varies, but how, geometrically, is it singular when $\mathbf{h}$ is degenerate? This problem is quite difficult, but must be solved before the true singular geometry of tensor singularities can be appreciated.

6.10   Appendix: Spherical harmonics in tensor bases

In this section we recall some elementary properties of the spherical harmonics $Y_l^m(\theta, \phi)$, and relate them to tensors with ranks 1 and 2 for $l = 1, 2$. Unlike the rest of this chapter, notation will be somewhat more conventional; the spin index m runs from $-l$ to l (rather than μ running from 0 to $2s = n$), and cartesian indices are x, y, z; integers are reserved for spin indices, and the axis of rotation is the z-axis. Square brackets are used for objects in a spin basis, round for objects in a cartesian basis.

Spherical harmonics are simply an appropriate form (linear if $l = 1$, quadratic if $l = 2$) of the unit polar vector $\mathbf{u}(\theta, \phi) = (\cos \phi \sin \theta, \sin \phi \sin \theta, \cos \theta)$; the tensors which make the form are a basis with respect to the inner product.

The spin 1 spherical harmonics Y_1^m correspond to the vectors

$$\mathbf{E}^1 = -\frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ i \\ 0 \end{bmatrix}, \qquad \mathbf{E}^0 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \qquad \mathbf{E}^{-1} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -i \\ 0 \end{bmatrix} \qquad (6.10.1)$$

and, with the spherical harmonics given their usual definition [Mac27], for $m = 1, 0, -1$,

$$Y_1^m(\theta, \phi) = \sqrt{\frac{3}{4\pi}} \mathbf{E}^m \cdot \mathbf{u}(\theta, \phi). \qquad (6.10.2)$$

The $\mathbf{E}^m$ form an orthonormal basis with respect to the usual inner product, and transform to the usual cartesian basis $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$ by the unitary transformations (cf (A.3.3))

$$\mathbf{M}_{1, s \to c} = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 & 0 & 1 \\ -i & 0 & -i \\ 0 & \sqrt{2} & 0 \end{pmatrix}, \qquad \mathbf{M}_{1, c \to s} = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 & i & 1 \\ 0 & 0 & \sqrt{2} \\ 1 & i & 0 \end{pmatrix}. \qquad (6.10.3)$$

Note that $\mathbf{E}^1$ is the negative of the two dimensional right-handed circular vector defined in (A.3.1); this is due to the conventions of spherical harmonics, constructed from raising and lowering operators (the "Condon-Shortley convention" [Alt86]).

For $l = 2$ the cartesian objects are traceless symmetric matrices, of the form (6.8.9), with inner product defined by (6.8.8). The matrices corresponding to the spherical harmonics Y_2^m are found to be

$$\mathbf{h}^{+2} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & i & 0 \\ i & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{h}^{+1} = -\frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & i \\ 1 & i & 0 \end{pmatrix}, \quad \mathbf{h}^0 = \sqrt{\frac{1}{3}} \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{pmatrix},$$

$$\mathbf{h}^{-1} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & -i \\ 1 & -i & 0 \end{pmatrix}, \quad \mathbf{h}^{-2} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -i & 0 \\ -i & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{6.10.4}$$

and, for $m = -2, \ldots, +2$,

$$Y_2^m = \sqrt{\frac{15}{16\pi}} \mathbf{u} \cdot \mathbf{h}^m \cdot \mathbf{u}. \tag{6.10.5}$$

A spin 2 state $\mathbf{h}$ is specified by five parameters (see (6.8.9)), which is written as a five-dimensional vector $\{U, V, X, Y, Z\}$, called cartesian because the basis states of this representation are orthogonal with respect to the inner product (6.8.8). The relevant five-dimensional unitary matrices transforming between cartesian and spherical bases are

$$\mathbf{M}_{2,\,s\to c} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & \sqrt{2} & 0 & 0 \\ 0 & -i & 0 & -i & 0 \\ 0 & -1 & 0 & 1 & 0 \\ i & 0 & 0 & 0 & -i \end{pmatrix}, \quad \mathbf{M}_{2,\,c\to s} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 0 & 0 & -i \\ 0 & 0 & i & -1 & 0 \\ 0 & \sqrt{2} & 0 & 0 & 0 \\ 0 & 0 & i & 1 & 0 \\ 1 & 0 & 0 & 0 & i \end{pmatrix}. \tag{6.10.6}$$

Note that, as with spherical harmonics, the only states totally orthogonal to the z-direction are those of $\pm m$.

Appendix A

The geometry of ellipses

'We then have first to understand - we must start with something - we first must know what an ellipse is.'

Richard Feynman, in *Feynman's lost lecture,*, eds. Goodstein and Goodstein, Vintage, 1997

Since much of the material in this thesis is dependent on ellipse geometry (either from the polarization ellipse of vector waves or the anisotropy ellipse of intensity describing the phase squeezing around a dislocation core) this appendix summarises the basic properties of ellipses in two and three dimensions that are used in the text.

A.1   Basic coordinate geometry

We shall start with coordinate geometry. An *ellipse* is defined to be the locus of points (up to rotation and translation) with cartesian coordinates (x, y) satisfying the equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \tag{A.1.1}$$

where a, b are real and $a > b \geq 0$. a is called the *major semiaxis* (or *semimajor axis*), and the points $(\pm a, 0)$ are the points on the ellipse furthest from the origin (and each other); b is the *minor semiaxis* (or *semiminor axis*), and $(0, \pm b)$ are the ellipse points closest to the origin.

Equation (A.1.1) generalises the equation for a circle radius r, centre 0 (ie $(x^2 + y^2)/r^2 = 1$), so the ellipse is a circle if $a = b$. If $b = 0$ then $y = 0$ always and the ellipse is a line of
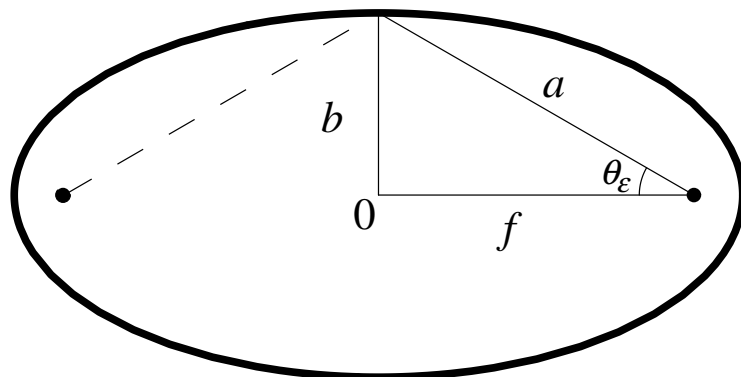
Figure A.1: Ellipse coordinate geometry, showing the pythagorean relationship between a, b, f and the eccentric angle θ_ε.

length $2a$ on the x-axis. The traditional construction of the ellipse, as the locus of points the sum L of whose distances from two points (*foci*) is always constant (easy to construct with two pegs and a piece of string). If the foci have coordinates $(\pm f, 0)$ (f is the *focal length*), then, as shown in figure (A.1), $L = 2a$, and at $(0, \pm b)$, there is a pythagorean relationship between a, b, f :

$$a^2 = b^2 + f^2. \tag{A.1.2}$$

In this triangle, the angle θ_ε opposite b is called the *eccentric angle* of the ellipse, and the *eccentricity* ε of the ellipse is defined to be

$$\varepsilon \equiv f/a = \cos\theta_\varepsilon = \sqrt{1 - \frac{b^2}{a^2}}. \tag{A.1.3}$$

ε has values between 0 and 1 - $\varepsilon = 0$ implies that $b = a$ and the ellipse is circular, whereas $\varepsilon = 1$ implies $b = 0$, and the ellipse is linear.

Different authors, particularly in ellipsometry (eg [Woa00, Bar85]) use other parameters to describe the degree of eccentricity of the ellipse, such as the *ellipticity* $b/a = \sin\theta_\varepsilon$, but here the eccentricity ε is used, which is 0 when the ellipse is circular (emphasising the similarity between C points in ellipse fields and zeros in scalar fields).

Writing equation (A.1.1) in terms of plane polar coordinates (R, ϕ) leads to a parametric equation for the ellipse,

$$R(\phi) = \frac{ab}{\sqrt{b^2 \cos^2\phi + a^2 \sin^2\phi}}. \tag{A.1.4}$$

If the ellipse is rotated such that its major semiaxis makes an angle ϕ_0 with the x-axis,
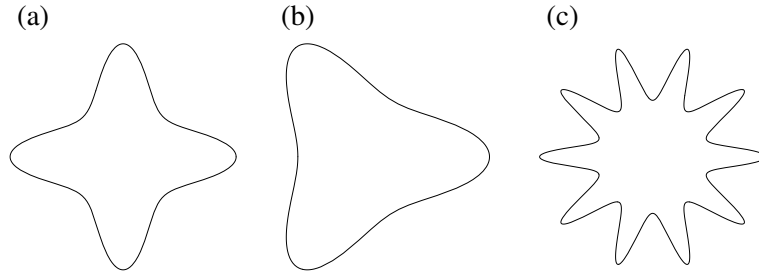
Figure A.2: Generalised ellipses of equation (A.1.7): (a) $n = 4$, (b) $n = 3$, (c) $n = 10$.

the equation is

$$R_{\text{rot}}(\phi) = \frac{ab}{\sqrt{b^2 \cos^2(\phi - \phi_0) + a^2 \sin^2(\phi - \phi_0)}}. \tag{A.1.5}$$

Equation (A.1.4) can be integrated to give the area of the ellipse,

$$\int_0^{2\pi} \mathrm{d}\phi \, R(\phi)^2 = \pi ab, \tag{A.1.6}$$

generalising the area of a circle, πr^2. Note that the symmetry of the ellipse means that $R(\phi) = R(-\phi) = R(\phi + \pi)$. Also note that the ellipse can be generalised to a family of closed curves with an eccentricity parameter $\varepsilon = \sqrt{1 - b^2/a^2}$, but with n lines of reflective symmetry (spaced at π/n - an ellipse has two at right angles), with equation

$$R_n(\phi) = \frac{ab}{\sqrt{b^2 \cos^2 \frac{n\phi}{2} + a^2 \sin^2 \frac{n\phi}{2}}}. \tag{A.1.7}$$

Some examples of generalised ellipses are shown in figure (A.2).

Another parameterisation of the ellipse is the locus traced out by the vector with cartesian coordinates

$$\mathbf{W}(\chi) = (a \cos \chi, \pm b \sin \chi) \tag{A.1.8}$$

as $0 \leq \chi \leq 2\pi$; such points satisfy equation (A.1.1), the sign giving the sense in which the ellipse is traced out, anticlockwise for $+$, clockwise for $-$. The angle parameter χ is called here the *phase angle*, since, up to a global rotation, this is the vector traced out by its real part as phase χ is varied, as appears in various places throughout this thesis (and is discussed in the following section). In the geometric context, χ is also called the *auxiliary angle*, since it is the (polar) angle on the *auxiliary circle* of radius $2a$ circumscribing the ellipse, as in figure (A.3). It is shown and discussed in section 2.3 that sectors of the ellipse with equal intervals of χ have equal areas; this is one of the two ways that angular
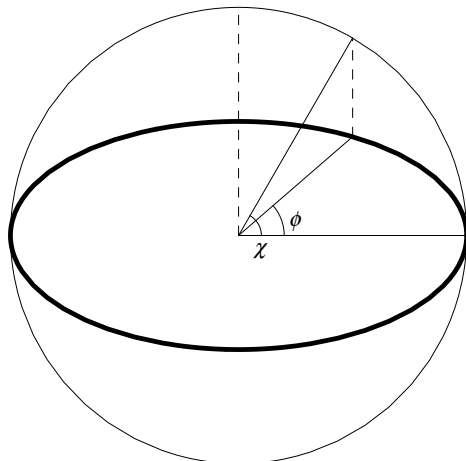
Figure A.3: The auxiliary circle, with phase angle χ and the corresponding polar angle ϕ.

momentum can be conserved on an elliptical orbit, the other of course being an inverse square central force directed towards a focus. The relationship between the two types of orbit is examined by [Nee97].

A.2   Ellipses and linear algebra

Equation (A.1.8) shows that ellipses may be easily described by vectors, as the image of the circle $\mathbf{U}(\chi) = (\cos\chi, \sin\chi)$ under a diagonal transformation

$$\mathbf{W}(\chi) = \begin{pmatrix} a & 0 \\ 0 & \pm b \end{pmatrix} \begin{pmatrix} \cos\chi \\ \sin\chi \end{pmatrix}. \tag{A.2.1}$$

This is in fact general for any real 2×2 real symmetric matrix $\mathbf{M}$; $\mathbf{MU}$ traces out an ellipse with semiaxes equal to the eigenvalues, the semiaxis directions given by the eigenvectors, and the sign of the determinant determining the sense of rotation of the vector. It also provides another way of finding the ellipse area πab. The hessian matrix of a real two dimensional function describes the Gauss curvature ellipse, as discussed in section 1.2. Throughout this section we shall normally be considering vectors in two dimensions, but most of the results generalise in an obvious way to three dimensions.

For a complex two dimensional vector ($\mathbf{X}, \mathbf{Y}$ real)

$$\mathbf{V} = \mathbf{X} + \mathrm{i}\mathbf{Y}, \tag{A.2.2}$$

the real part traces out an ellipse as the phase is varied, ie

$$\mathbf{W}(\chi) = \operatorname{Re} \exp(-\mathrm{i}\chi)\mathbf{V} = \mathbf{X}\cos\chi + \mathrm{i}\mathbf{Y}\sin\chi, \tag{A.2.3}$$

generalising equation (A.1.8), since this holds even if $\mathbf{X}, \mathbf{Y}$ are not orthogonal. If $\varepsilon > 0$, there is always a phase χ_0, where

$$
\begin{aligned}
\mathbf{V}_0 &= \exp(-\mathrm{i}\chi_0)\mathbf{V} \\
&= (\mathbf{X}\cos\chi_0 + \mathbf{Y}\sin\chi_0) + \mathrm{i}(\mathbf{Y}\cos\chi_0 - \mathbf{X}\sin\chi_0) \\
&= \mathbf{X}_0 + \mathrm{i}\mathbf{Y}_0
\end{aligned}
\tag{A.2.4}
$$

such that $\mathbf{X}_0 \cdot \mathbf{Y}_0 = 0$, which occurs when

$$2\mathbf{X} \cdot \mathbf{Y} \cos 2\chi_0 = (X^2 - Y^2)\sin 2\chi_0, \tag{A.2.5}$$

that is, such that

$$\tan 2\chi_0 = \frac{2\mathbf{X} \cdot \mathbf{Y}}{X^2 - Y^2}. \tag{A.2.6}$$

χ_0 shall be referred to as the *rectifying phase* of the ellipse; it measures the phase discrepancy between $\mathbf{X}, \mathbf{Y}$ and the orthogonal semiaxis frame of the ellipse. It is singular (not defined) if $\varepsilon = 0$, (ie $\mathbf{X} \cdot \mathbf{Y} = 0$, $X^2 - Y^2 = 0$,) and C points are therefore rectifying phase singularities.

The original $\mathbf{X}, \mathbf{Y}$ are interpreted geometrically as conjugate semiradii of the ellipse, that is, the tangent to the ellipse at $\mathbf{X}$ is parallel to $\mathbf{Y}$, and vice versa. This is easily seen from equation (A.2.3), since the tangent to the ellipse at phase angle χ is in the direction

$$\mathbf{W}'(\chi) = \frac{\mathrm{d}\mathbf{W}}{\mathrm{d}\chi} = -\mathbf{X}\sin\chi + \mathbf{Y}\cos\chi \tag{A.2.7}$$

so, at $\mathbf{X}$, $\chi = 0$ and $\mathbf{W}' = \mathbf{Y}$ as required (similarly at $\mathbf{Y}$, $\chi = \pi/2$ and $\mathbf{W}' = -\mathbf{X}$). In fact, any two phase vectors $\mathbf{W}(\chi), \mathbf{W}(\chi + \pi/2)$ are conjugate semiradii, and are also phase conjugates in the sense of section 2.1.

When the phase angle is not important, the ellipse can be considered as a contour of a quadratic form, derived from the real symmetric matrix $\mathbf{M} = \left(\begin{smallmatrix} 1/a^2 & 0 \\ 0 & 1/b^2 \end{smallmatrix} \right)$ in (A.1.1), and more generally by rotating this in the plane. This interpretation is relevant when, as with the hessian mentioned above, one wants to consider the contour shape of a surface in two dimensions, such as ρ^2 from (2.3.1), which is, in general,

$$
\begin{aligned}
\rho^2 &= \text{constant} + (\mathbf{R} \cdot \nabla\xi)^2 + (\mathbf{R} \cdot \nabla\eta)^2 \\
&= \text{constant} + \mathbf{R}^T (\nabla\xi \otimes \nabla\xi + \nabla\eta \otimes \nabla\eta)\mathbf{R}.
\end{aligned}
\tag{A.2.8}
$$

The quadratic form ellipse has major and minor semiaxes interchanged from the phase ellipse traced out by $\mathbf{W}$ in equation (A.2.1), and equation (A.2.8) makes it clear why: in (A.2.1), $\mathbf{X}, \mathbf{Y}$, are vectors with units of length, whereas in (A.2.8) $\nabla\xi, \nabla\eta$ are forms (contour normals) with dimensions of inverse length (see for example, [Sch80]).

Equation (A.2.8) gives a way of constructing an orthogonal basis in real n-dimensional space (more democratic than the usual Gram-Schmidt orthogonalisation procedure), given n linearly independent real vectors $\mathbf{X}_1, \ldots, \mathbf{X}_n$. The real, symmetric $n \times n$ matrix

$$\mathbf{M}_n = \sum_{j=1}^{n} \mathbf{X}_j \otimes \mathbf{X}_j \tag{A.2.9}$$

gives the locus (when taken as an n-dimensional quadratic form) of an n-dimensional ellipsoid on which $\mathbf{X}_1, \ldots \mathbf{X}_n$ lie; the eigenvectors of $\mathbf{M}_n$ provide an orthogonal basis (provided there are no degeneracies), and if $n = 2$, these are precisely $\mathbf{X}_0, \mathbf{Y}_0$ of equation (A.2.4).

This result can also be used to find the mutual perpendicular, $\mathbf{Z}$, to a set of $n - 1$ linearly independent vectors $\mathbf{X}_1, \ldots \mathbf{X}_{n-1}$: it is the null eigendirection of the matrix

$$\mathbf{M}_{n,n-1} = \sum_{j=1}^{n-1} \mathbf{X}_j \otimes \mathbf{X}_j, \tag{A.2.10}$$

and is equal to the vector given by the generalised cross product (where components of the vectors $\mathbf{X}_j$ are written as superscripts, and $\varepsilon_{ab\ldots p}$ is the general antisymmetric symbol)

$$Z^a = \varepsilon_{ab\ldots p} X_1^b \ldots X_{n-1}^p. \tag{A.2.11}$$

In particular, if $n = 3$, and $\mathbf{p} + i\mathbf{q}$ is a complex three dimensional vector (see equation (4.3.1)), the flat ellipsoid (ie the ellipse) of the matrix

$$\mathbf{M}_{3,2} = \mathbf{p} \otimes \mathbf{p} + \mathbf{q} \otimes \mathbf{q} \tag{A.2.12}$$

has normal in direction $\mathbf{p} \wedge \mathbf{q}$.

The eigenvalues of $\mathbf{M}_2$ (and nonzero eigenvalues for $\mathbf{M}_{3,2}$, with appropriate replacements of symbols), are given by

$$\mu_\pm = \frac{1}{2}\left(X^2 + Y^2 \pm \sqrt{(X^2 + Y^2)^2 - 4(\mathbf{X} \wedge \mathbf{Y})^2}\right) \tag{A.2.13}$$

and the eccentricity of the ellipse is

$$
\begin{aligned}
\varepsilon^2 &= 1 - \frac{\mu_-}{\mu_+} \\
&= \frac{1}{2(\mathbf{X} \wedge \mathbf{Y})^2} \left(X^2 + Y^2 - \sqrt{(X^2 + Y^2)^2 - 4(\mathbf{X} \wedge \mathbf{Y})^2} \right) \\
&\quad \times \sqrt{(X^2 + Y^2)^2 - 4(\mathbf{X} \wedge \mathbf{Y})^2}.
\end{aligned}
\tag{A.2.14}
$$

The eigenvectors corresponding to $\mu_\pm$ can be found using χ_0 from equations (A.2.4), (A.2.6). The expressions are messy and complicated, and are not given here.

A.3 Ellipses and the complex projective line

Another way of encoding the information of a two dimensional phase ellipse (as in equation (A.2.3)) is by using a circular instead of cartesian basis, with right and left handed circular vectors are defined by

$$
\mathbf{e}_\mathrm{R} \equiv \frac{1}{\sqrt{2}} (\mathbf{e}_x + i\mathbf{e}_y),
\tag{A.3.1}
$$

$$
\mathbf{e}_\mathrm{L} \equiv \frac{1}{\sqrt{2}} (\mathbf{e}_x - i\mathbf{e}_y).
\tag{A.3.2}
$$

Note that the definition of $\mathbf{e}_\mathrm{R}$ used here is -1 times that used in chapter 6 (which comes from symmetries of spherical harmonics). The basis transformation matrices are (in obvious notation)

$$
\mathbf{M}_{+\to\circ} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix}, \qquad \mathbf{M}_{\circ\to+} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & i \\ 1 & -i \end{pmatrix}.
\tag{A.3.3}
$$

Thus $\mathbf{V}$ in equation (A.2.2) may be rewritten in circular coordinates $\alpha_+ \mathbf{e}_\mathrm{R} + \alpha_- \mathbf{e}_\mathrm{L}$ ($\alpha_\pm$ complex), where

$$
\alpha_\pm = \frac{1}{\sqrt{2}} (V_x \mp i V_y).
\tag{A.3.4}
$$

In fact, $\begin{bmatrix} \alpha_+ \\ \alpha_- \end{bmatrix}$ is an elementary spinor, discussed extensively in chapter 6, where different notation is used because different conventions are. All complex vectors shall now be assumed normalised (that is, $|\mathbf{V}| = \sqrt{|\alpha_+|^2 + |\alpha_-|^2} = 1$). The (normalised) ellipse is traced out in the complex plane by associating $\mathbf{e}_\mathrm{R}$ with $\exp(i\chi)$ and $\mathbf{e}_\mathrm{L}$ with $\exp(-i\chi)$, that is the locus of complex numbers

$$
\alpha_+ \exp(i\chi) + \alpha_- \exp(-i\chi)
\tag{A.3.5}
$$

which trace out an ellipse about the origin. It can be shown [Nee97] that

$$
\begin{aligned}
a &= \frac{1}{\sqrt{2}}(|\alpha_+| + |\alpha_-|), \\
b &= \frac{1}{\sqrt{2}}(|\alpha_+| - |\alpha_-|), \\
\varepsilon^2 &= \frac{4|\alpha_+||\alpha_-|}{(|\alpha_+| + |\alpha_-|)^2}, \\
\phi_0 &= \frac{1}{2}(\arg \alpha_+ - \arg \alpha_-),
\end{aligned}
\tag{A.3.6}
$$

where the sign of b, as before, gives the sense of rotation of the ellipse, and ϕ_0 is as in equation (A.1.5). Four real parameters describe an ellipse, the real and imaginary parts each of $\alpha_\pm$: these four freedoms correspond to overall size (equivalent to normalising), angular orientation ϕ_0, eccentricity ε, and phase (equivalent to χ_0, the position on the ellipse when $\chi = 0$ in (A.3.5)). If we are only interested in the orientation and eccentricity, the phase (which is the overall phase of the spinor $\begin{bmatrix} \alpha_+ \\ \alpha_- \end{bmatrix}$) is redundant, and these ellipse geometry parameters can be described by one complex number (possibly ∞):

$$
z = \frac{\alpha_-}{\alpha_+},
\tag{A.3.7}
$$

Note that equation (A.3.7) gives a bijective correspondence between all possible normalised ellipses (of both handednesses) and the complex plane (including ∞). $z = 0$ corresponds to a righthanded circle; $z = \infty$ corresponds to a lefthanded circle. The circle in the complex plane $|z| = 1$ corresponds to the linear ellipse with $\varepsilon = 1$, and argument giving the possible orientations (mod π).

Since the complex coordinates α_+, α_- can be multiplied through by any complex number without changing the ratio α_-/α_+, the space of all such pairs is called the *complex projective space* of dimension 1, or complex projective line (since it has only one complex freedom). This generalises to n dimensional complex projective space when there are $n+1$ complex numbers whose ratios are invariant, and these are related to the various spaces of spinors, as discussed in chapter 6. There are well-known connections between two dimensional complex spaces and three dimensional real spaces (mainly due to the fact that the universal covering group of the 3D rotation group SO(3) is the 2D unitary group SU(2), [Nee97]). Advantage is taken of this to give a further geometrical description of ellipses in the following section.

A.4   The Poincaré sphere and Stokes parameters

Given that the complex projective line is the complex plane including ∞, it is natural to consider stereographic projection of z in equation (A.3.7) onto the unit sphere (from the south pole), with polar angles (α, β), where

$$\alpha = 2 \arctan |z|, \qquad \beta = \arg z, \qquad (A.4.1)$$

and

$$z = \tan \alpha/2 \, \exp(\mathrm{i}\beta). \qquad (A.4.2)$$

This sphere is called the *Riemann sphere* in the theory of the complex projective line [Nee97, Fra97], but when it parameterises (polarization) ellipses, it is called the *Poincaré sphere* [Poi92]. Ellipses for the remainder shall be described using the language of polarization. The Poincaré sphere is in 1-1 correspondence with the different possible states of polarization ellipses in the plane (up to overall size and phase, as described previously). By equations (A.3.6), (A.3.7), (A.4.1), the Poincaré sphere azimuth β is clearly $\phi_0/2$; since an ellipse is symmetric with respect to a rotation by π, there is no redundancy here (although it means that the space of the Poincaré sphere is not real space). The colatitude angle α is related to the eccentricity by the same equations; if $0 \leq \alpha < \pi/2$ (the 'northern hemisphere'), polarization is right handed, with circular polarization at the north pole. If $\pi/2 < \alpha \leq \pi$ (the 'southern hemisphere'), the polarization is left handed, with circular polarization at the south pole. Polarization is linear on the equator ($\alpha = \pi/2$), and reflection in the equatorial plane ($\alpha \to \pi - \alpha$) reverses the handedness of the polarization, but does not otherwise change the ellipse. α and ε are related directly by

$$\varepsilon = \frac{4 \tan \alpha/2}{(1 + \tan \alpha/2)^2}. \qquad (A.4.3)$$

A visualisation of the Poincaré sphere is shown in figure (A.4).

The spinor $[\alpha_+, \alpha_-] = [\cos \alpha/2 \exp(-\mathrm{i}\beta/2), \sin \alpha/2 \exp(\mathrm{i}\beta/2)]$ has a convenient representation in terms of the Riemann-Poincaré sphere, via a remarkable result attributed to Archimedes (implicit whenever one integrates over the sphere, but stated explicitly here). The spinor space can geometrically be represented by a cylinder, of height $h = |\alpha_+| - 1/2$ (recalling that the spinor is normalised), and azimuth $\beta = 1/2(\arg \alpha_+ - \arg \alpha_-)$ (we are not concerned about the phase singularities at $h = \pm 1/2$), which circumscribes the unit Riemann sphere. Archimedes' result is that any strip of the cylinder, of width Δh around
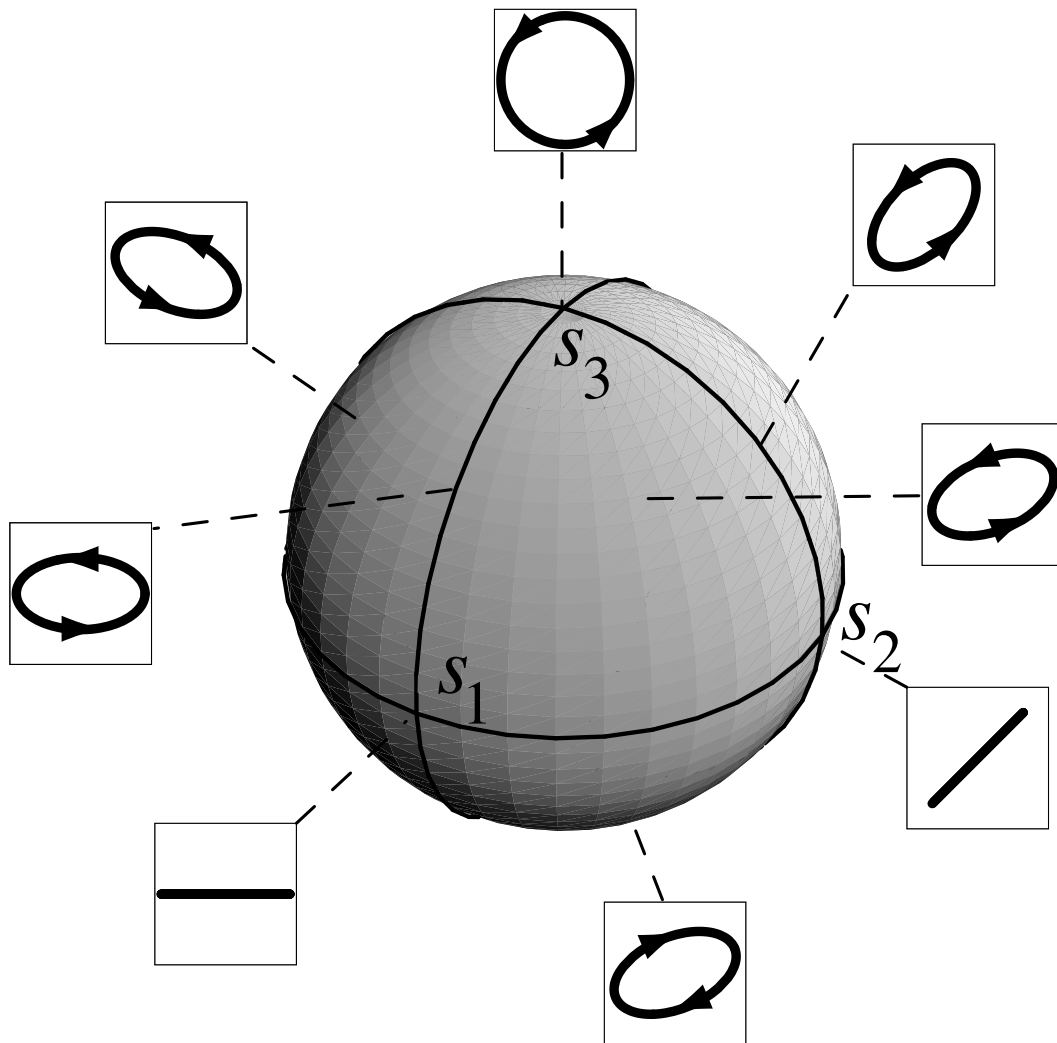
Figure A.4: The Poincaré sphere, with cartesian axes (s_1, s_2, s_3), indicating the polar parameters for certain ellipses.

height h has equal area on the cylinder and on the sphere, as $h = \cos \alpha$, and the measure of $\cos \alpha$ is uniform on the sphere. Thus the natural Hilbert space measure (where α_+, α_- are identically and independently distributed) corresponds to the natural geometric measure on the Poincaré sphere; that is, statistically, sets on the Poincaré sphere of equal area are equally likely, a fact used in the vector wave statistics of chapter 4.

The geometric interpretation of the Poincaré sphere is also implied (via the aforementioned connection between SU(2) and SO(3)) by the fact that, if $\mathbf{s} = (s_1, s_2, s_3)$ is a (unit) vector on the surface of the sphere, and $\boldsymbol{\sigma}$ is the 3-vector of Pauli spin matrices

$$(\sigma_1, \sigma_2, \sigma_3) = \left(\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \right), \tag{A.4.4}$$

then a 2×2 hermitian matrix $\mathbf{H}(\mathbf{s})$ is defined

$$\mathbf{H}(\mathbf{s}) = \frac{1}{2} \mathbf{s} \cdot \boldsymbol{\sigma} = \frac{1}{2} \begin{bmatrix} s_3 & s_1 - \mathrm{i} s_2 \\ s_1 + \mathrm{i} s_2 & -s_3 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} \cos \alpha & \sin \alpha \exp(-\mathrm{i}\beta) \\ \sin \alpha \exp(\mathrm{i}\beta) & -\cos \alpha \end{bmatrix}. \tag{A.4.5}$$

If the spinor $\begin{bmatrix} \alpha_+ \\ \alpha_- \end{bmatrix}$ represents a quantum mechanical state, then $\mathbf{H}$ is the hamiltonian for the system (see [Ber87] for more details). $\mathbf{H}$ has eigenvalues $\pm 1/2$, with $\begin{bmatrix} \alpha_+ \\ \alpha_- \end{bmatrix}$ the eigenspinor for $+1/2$. More details of this geometry are discussed in chapter 6.

The cartesian coordinates of $\mathbf{s}$ can, as usual, be written in terms of the polar angles α, β, and the complex number z, of equation (A.3.7) may be rewritten

$$z = \frac{s_1 + \mathrm{i} s_2}{1 + s_3}. \tag{A.4.6}$$

In terms of the original (not normalised) vector $\mathbf{V} = \mathbf{X} + \mathrm{i}\mathbf{Y}$ the four *Stokes parameters* S_0, S_1, S_2, S_3 are defined

$$
\begin{aligned}
S_0 &= |\mathbf{V}|^2 \\
S_1 &= |V_x|^2 - |V_y|^2 = X_x^2 - X_y^2 + Y_x^2 - Y_y^2 \\
S_2 &= V_x^* V_y + V_y^* V_x = 2(X_x X_y + Y_x Y_y) \\
S_3 &= -\mathrm{i}(V_x^* V_y - V_y^* V_x) = 2(X_x Y_y - Y_x X_y) = 2\mathbf{X} \wedge \mathbf{Y}
\end{aligned}
\tag{A.4.7}
$$

and the components of $\mathbf{s}$ are the *normalised Stokes parameters*,

$$s_i = \frac{S_i}{S_0}, \qquad i = 1, 2, 3. \tag{A.4.8}$$

The vector $\mathbf{s}$ is called the *Stokes vector*. The eccentricity (A.2.14) can be rewritten in terms of the Stokes parameters

$$
\begin{aligned}
\varepsilon^2 &= \frac{2}{S_3^2}\left(S_0 - \sqrt{S_0^2 - S_3^2}\right)\sqrt{S_0^2 - S_3^2} \\
&= \frac{2}{s_3^2}\left(1 - \sqrt{1 - s_3^2}\right)\sqrt{1 - s_3^2},
\end{aligned} \tag{A.4.9}
$$

implying that

$$
s_3 = \pm\frac{2\sqrt{1 - \varepsilon^2}}{2 - \varepsilon^2}. \tag{A.4.10}
$$

The the angle of the major semiaxis ϕ_0 is

$$
\phi_0 = \frac{1}{2}\arctan\frac{S_2}{S_1}. \tag{A.4.11}
$$

The Stokes parameters are phase (gauge) invariant; multiplying the complex vector $\mathbf{V}$ by any phase does not change the Stokes parameters (just as it does not change the ellipse).

Note that the Poincaré sphere is only good for describing the polarization of vectors and ellipses in two dimensions (eg paraxial optical fields). The correct analogue of the Poincaré sphere in three dimensions is the Majorana sphere $\mathcal{M}_2$, described in relation to the Poincaré sphere in chapter 6.

Bibliography

[AA87]   Y. Aharonov and J.S. Anandan. Phase change during a cyclic quantum evolution. *Phys.Rev.Lett.*, **58**:1593–1596, 1987.

[AB59]   Y. Aharonov and D. Bohm. Significance of electromagnetic potentials in the quantum theory. *Phys.Rev.*, **115**:485–491, 1959.

[ABSW92]   L. Allen, M. Beijersbergen, R.J.C. Spreeuw, and J.P. Woerdman. Orbital angular momentum of light and the transformation of Laguerre-Gaussian laser modes. *Phys.Rev.A*, **45**:8185–8189, 1992.

[Ada94]   C.C. Adams. *The knot book*. Freeman, San Francisco, CA, 1994.

[Adl81]   R.J. Adler. *The geometry of random fields*. Wiley, 1981.

[Alt86]   S.L. Altmann. *Rotations, quaternions, and double groups*. Oxford University Press, 1986.

[APB99]   L. Allen, M.J. Padgett, and M. Babiker. The orbital angular momentum of light. *Prog.Opt.*, **39**:291–372, 1999.

[Arn86]   V.I. Arnold. *Catastrophe theory*. Springer, 1986.

[Arn89]   V.I. Arnold. *Mathematical methods of classical mechanics*. Springer, 2nd edition, 1989.

[AS65]   M. Abramowitz and I.A. Stegun, editors. *Handbook of Mathematical Functions*. Dover, 1965.

[Ati01]   M.F. Atiyah. The geometry of classical particles. In S.T. Yau, editor, *Surveys in Differential Geometry*. International Press, Cambridge, MA, 2001.

[Bar47]   V. Bargmann. Irreducible unitary representations of the Lorentz group. *Ann.Math.*, **48**:568–640, 1947.

[Bar85]   R. Barakat. The statistical properties of partially polarized light. *Opt.Act.*, **32**:295–312, 1985.

[Bar87]   R. Barakat. Statistics of the Stokes parameters. *J.Opt.Soc.Am.A*, **4**:1256–1263, 1987.

[BBBBS00]   I. Bialynicki-Birula, S. Bialynicka-Birula, and C. Sliwa. Motion of vortex lines in quantum mechanics. *Phys.Rev.A*, **61**(032110), 2000.

[BCL$^+$80]   M.V. Berry, R.G. Chambers, M.D. Large, C. Upstill, and J.C. Walmsley. Wavefront dislocations in the Aharonov-Bohm effect and its water-wave analogue. *Eur.J.Phys.*, **1**:154–162, 1980.

[BD00]   M.V. Berry and M.R. Dennis. Phase singularities in isotropic random waves. *Proc.R.Soc.Lond.A*, **456**:2059–2079, 2000. (errata **456**:3059).

[BD01a]   M.V. Berry and M.R. Dennis. Knotted and linked singularities in monochromatic waves. *Proc.R.Soc.Lond.A*, **457**:2251–2263, 2001.

[BD01b]   M.V. Berry and M.R. Dennis. Knotting and unknotting of phase singularities: Helmholtz waves, paraxial waves and waves in 2+1 dimensions. *J.Phys.A:Math.Gen.*, **34**:8877–8888, 2001.

[BD01c]   M.V. Berry and M.R. Dennis. Polarization singularities in isotropic random vector waves. *Proc.R.Soc.Lond.A*, **457**:141–155, 2001.

[Bea79]   A.F. Beardon. *Complex Analysis: The argument principle in analysis and topology.* Wiley, 1979.

[Ber77]   M.V. Berry. Regular and irregular semiclassical wavefunctions. *J.Phys.A:Math.Gen.*, **10**:2083–2091, 1977.

[Ber78]   M.V. Berry. Disruption of wavefronts: statistics of dislocations in incoherent Gaussian random waves. *J.Phys.A:Math.Gen.*, **11**:27–37, 1978.

[Ber80]   M.V. Berry. Some geometric aspects of wave motion: wavefront dislocations, diffraction catastrophes, diffractals. In R. Osserman and A. Weinstein, editors, *Geometry of the Laplace operator*, Proc.Symp.App.Maths, pages 13–28. AMS, 1980.

[Ber81]   M.V. Berry. Singularities in waves and rays. In R. Balian, M. Kléman, and J.-P. Poirier, editors, *Les Houches Session XXV - Physics of Defects*, pages 453–543. North-Holland, Amsterdam, 1981.

[Ber84]   M.V. Berry. Quantal phase factors accompanying adiabatic changes. *Proc.R.Soc.Lond.A*, **392**:45–57, 1984.

[Ber87]   M.V. Berry. The adiabatic phase and Pancharatnam's phase for polarized light. *J.Mod.Opt.*, **34**:1401–1407, 1987.

[Ber91a]   M.V. Berry. Bristol anholonomy calendar. In R.G. Chambers, J.E. Enderby, A. Keller, A. Lang, and J.W. Steeds, editors, *Sir Charles Frank OBE: an eightieth birthday tribute*, pages 207–219. Adam Hilger, Bristol, 1991.

[Ber91b]   M.V. Berry. Wave geometry: a plurality of singularities. In J.S. Anandan, editor, *Quantum Coherence*, pages 92–98. World Scientific, Singapore, 1991.

[Ber92]   M.V. Berry. Waves, wavefronts and phase: a picture book of cusps. In H. Bok and H.A. Ferwerda, editors, *Huygens' principle 1690-1990; theory and applications*, pages 97–111. North-Holland, Amsterdam, 1992.

[Ber94a]   M.V. Berry. Asymptotics, singularities and the reduction of theories. In D. Prawitz, B. Skyrms, and D. Westerståhl, editors, *Logic, Methodology and Philosophy of Science IX*, pages 597–607. Elsevier, 1994.

[Ber94b]   M.V. Berry. Evanescent and real waves in quantum billiards and Gaussian beams. *J.Phys.A:Math.Gen.*, **27**:L391–L398, 1994.

[Ber94c]   M.V. Berry. Pancharatnam, virtuoso of the Poincaré sphere: an appreciation. *Curr.Sci.*, **67**:220–223, 1994.

[Ber98]   M.V. Berry. Much ado about nothing: optical dislocation lines (phase singularities, zeros, vortices...). In M.S. Soskin, editor, *International Conference on Singular Optics*, volume **3487**, pages 1–15. SPIE, 1998.

[Ber01a]   M.V. Berry. Geometry of phase and polarization singularities, illustrated
             by edge diffraction and the tides. In M.S. Soskin and M.V. Vasnetsov,
             editors, *Singular Optics (Optical Vortices): Fundamentals and Applications*,
             volume **4403**, pages 1–12. SPIE, 2001.

[Ber01b]   M.V. Berry. Knotted zeros in the quantum states for hydrogen. *Found.Phys*,
             31:659–667, 2001.

[BF71]   F. Brock Fuller. The writhing number of a space curve. *Proc.N.A.S. U.S.A.*,
             68:815–819, 1971.

[BH77]   M.V. Berry and J.H. Hannay. Umbilic points on a Gaussian random surface.
             J.Phys.A:Math.Gen., **10**:2083–2091, 1977.

[BNW79]   M.V. Berry, J.F. Nye, and F.J. Wright. The elliptic umbilic diffraction
             catastrophe. *Phil.Trans.R.Soc.A*, **291**:453–484, 1979.

[Bou60]   R.C. Bourret. Coherence properties of blackbody radiation. *Nuov.Cim*,
             18:347–356, 1960.

[Boy68]   C.B. Boyer. *A history of mathematics*. Wiley, 1968.

[BPSS99]   K.-F. Berggren, K.N. Pichigin, A.F. Sadreev, and A. Starikov. Signatures
             of quantum chaos in the nodal points and streamlines in electron transport
             through billiards. *JETP Lett.*, **70**:403–409, 1999.

[BR86]   M.V. Berry and M. Robnik. Quantum states without time-reversal sym-
             metry: wavefront dislocations in a non-integrable Aharonov-Bohm billiard.
             J.Phys.A:Math.Gen., **19**:1365–1372, 1986.

[Bre47]   D. Brewster. On the polarization of the atmosphere. *Phil.Mag.*, **31**:444–454,
             1847.

[Bro95]   C. Brosseau. Statistics of the normalized Stokes parameters for a Gaussian
             stochastic plane wave field. *App.Opt.*, **34**:4788–4793, 1995.

[Bro98]   C. Brosseau. *Fundamentals of polarized light: A statistical optics approach*.
             Wiley, 1998.

[BU80]   M.V. Berry and C. Upstill. Catastrophe optics: Morphologies of caustics and their diffraction patterns. *Prog.Opt.*, **18**:257–346, 1980.

[BW48]   V. Bargmann and E.P. Wigner. Group theoretical discussion of relativistic wave equations. *Proc.N.A.S. U.S.A.*, **34**:211–223, 1948.

[BW59]   M. Born and E. Wolf. *Principles of Optics*. Pergamon, Oxford, 1959.

[BZM$^+$81] N.B. Baranova, B.Y. Zel'dovitch, A.V. Mamaev, N. Pilipetskii, and V.V. Shkukov. Dislocations of the wavefront of a speckle-inhomogeneous pattern (theory and experiment). *JETP Lett.*, **33**:195–199, 1981.

[Car66]   E. Cartan. *The theory of spinors*. MIT Press, Cambridge, MA, 1966.

[Car99]   D.E. Cartwright. *Tides: A scientific history*. Cambridge University Press, 1999.

[CBR98]   P. Cromwell, E. Beltrami, and M. Rampicini. The Borromean rings. *Math.Int.*, **20**:53–62, 1998.

[CG82]   L. Campbell and W. Garnett. *The life of James Clerk Maxwell*. Macmillan, London, 1882.

[Cha50]   S. Chandrasekhar. *Radiative transfer*. Oxford University Press, 1950.

[CLH56]   D.E. Cartwright and M.S. Longuet-Higgins. The statistical distribution of the maxima of a random function. *Proc.R.Soc.Lond.A*, **237**:212–232, 1956.

[Cor53]   E.M. Corson. *Introduction to tensors, spinors and relativistic wave-equations*. Blackie and Son, Glasgow, 1953.

[Dai76]   J.C. Dainty. The statistics of speckle patterns. *Prog.Opt.*, **14**:1–46, 1976.

[Dar96]   G. Darboux. *Leçons sur la théorie generale des surfaces et les applications géométriques du calcul infinitésimal*, volume 4. Gauthiers-Villars, Paris, 1896.

[dC76]   M.P. do Carmo. *Differential geometry of curves and surfaces*. Prentice-Hall, Englewood Cliffs, NJ, 1976.

[Den00]   M.R. Dennis. The topology of twisted phase singularities, 2000. (Unpublished).

[Den01a]   M.R. Dennis. Local properties and statistics of phase singularities in generic wavefields. In M.S. Soskin and M.V. Vasnetsov, editors, *Singular Optics (Optical Vortices): Fundamentals and Applications*, volume **4403**, pages 13–23. SPIE, 2001.

[Den01b]   M.R. Dennis. Phase critical point densities in planar isotropic random waves. *J.Phys.A:Math.Gen.*, **34**:L297–L303, 2001.

[Dir31]   P.A.M. Dirac. Quantised singularities in the electromagnetic field. *Proc.R.Soc.Lond.A*, **133**:60–72, 1931.

[DMJE87]   J. Durnin, J.J. Miceli Jr, and J.H. Eberly. Diffraction-free beams. *Phys.Rev.Lett.*, **58**:1499–1501, 1987.

[Dur87]   J. Durnin. Exact solutions for nondiffracting beams. I. The scalar theory. *J.Opt.Soc.Am.A*, **4**:651–654, 1987.

[Eco98]   U. Eco. *The island of the day before.* Vintage, 1998.

[EH10a]   A. Einstein and L. Hopf. On a theorem of the probability calculus and its application to the theory of radiation. *Ann.Phys.*, **33**:1096–1104, 1910.

[EH10b]   A. Einstein and L. Hopf. Statistical investigation of a resonator's motion in a radiation field. *Ann.Phys.*, **33**:1105–1115, 1910.

[Eis60]   L.P. Eisenhart. *A treatise on the differential geometry of curves and surfaces.* Dover, New York, 1960.

[FB00]   I. Freund and A. Belenkiy. Higher-order extrema in two-dimensional wave fields. *Opt.Commun.*, **17**:434–446, 2000.

[Fel50]   W. Feller. *An introduction to probability theory and its applications*, volume I. Wiley, New York, 3rd edition, 1950.

[Fey97]   R.P. Feynman. The motion of planets around the sun. In D.L. Goodstein and J.R. Goodstein, editors, *Feynman's lost lecture.* Vintage, 1997.

[FF97]   I. Freund and V. Freilikher.   Parameterization of anisotropic vortices. *J.Opt.Soc.Am.A*, **14**:1902–1910, 1997.

[FG82]   P.A. Firby and C.F. Gardiner. *Surface topology.* Ellis Horwood, Chichester, 1982.

[FG87]   J. Fauvel and J. Gray, editors.   *The History of Mathematics: A reader.* Macmillan, 1987.

[FK01]   I. Freund and D. Kessler. Critical point trajectory bundles in singulaw wave fields. *Opt.Commun.*, **187**:71–90, 2001.

[FLS63a]   R.P. Feynman, R.B. Leighton, and M. Sands.   *The Feynman Lectures on Physics*, volume 2. Addison-Wesley, 1963.

[FLS63b]   R.P. Feynman, R.B. Leighton, and M. Sands.   *The Feynman Lectures on Physics*, volume 3. Addison-Wesley, 1963.

[FMW99]   R.P. Feynman, F.B. Morinigo, and W.G. Wagner.   *Feynman Lectures on Gravitation.* Penguin, 1999.

[Fra58]   F.C. Frank. On the theory of liquid crystals. *Farad.Soc.Disc.*, **25**:19–28, 1958.

[Fra88]   F.C. Frank. Orientation mapping. *Metall.Trans.A*, **19**:1141–1149, 1988.

[Fra97]   T. Frankel. *The geometry of physics: an introduction.* Cambridge University Press, 1997.

[Fre94]   I. Freund.   Optical vortices in Gaussian random wave-fields - statistical probability densities. *J.Opt.Soc.Am.A*, **11**:1644–1652, 1994.

[Fre95]   I. Freund.   Saddles, singularities and extrema in random phase fields. *Phys.Rev.E*, **52**:2346–2360, 1995.

[Fre96]   I. Freund.   Intensity critical point correlation functions in random wave fields. *Opt.Commun.*, **128**:315–324, 1996.

[Fre97]   I. Freund.   Critical-point level-crossing geometry in random wave fields. *J.Opt.Soc.Am.A*, **14**:1911–1927, 1997.

[Fre98a]   I. Freund. '1001' correlations in random wave fields. *Waves Rand.Med.*, **8**:119–158, 1998.

[Fre98b]   I. Freund. Phase correlations at neighboring intensity critical points in Gaussian random wave fields. *App.Opt.*, **37**:7560–7567, 1998.

[Fre00]   I. Freund. Optical vortex trajectories. *Opt.Commun.*, **181**:19–33, 2000.

[Fre01]   I. Freund. Vortex Flowers. *Opt.Commun.*, **196**:63–76, 2001.

[Fri95]   U. Frisch. *Turbulence.* Cambridge University Press, 1995.

[FS94]   I. Freund and N. Shvartsman. Wave-field phase singularities: the sign principle. *Phys.Rev.A*, **50**:5164–5172, 1994.

[FW98]   I. Freund and M. Wilkinson. Critical-point screening in random wave fields. *J.Opt.Soc.Am.A*, **15**:2892–2902, 1998.

[Gab46]   D. Gabor. Theory of communication. *J.Inst.Elec.Eng.*, **93**:429–447, 1946.

[GHMRW01] J. Gal, G. Horvath, V.B. Meyer-Rochow, and R. Wehner. Polarization patterns of the summer sky and its neutral points measured by full-sky imaging polarimetry in Finnish Lapland north of the Arctic Circle. *Proc.R.Soc.Lond.A*, **457**:1385–1399, 2001.

[Goo75]   J.W. Goodman. Statistical properties of laser speckle patterns. In J.C. Dainty, editor, *Laser speckle and related phenomena*, pages 9–75. Springer-Verlag, 1975.

[Goo85]   J.W. Goodman. *Statistical Optics.* Wiley, 1985.

[Gra93]   A. Gray. *Modern differential geometry of curves and surfaces with Mathematica.* CRC Press, 1993.

[Gri87]   D. Griffiths. *Introduction to elementary particles.* Wiley, 1987.

[Haj85]   J.V. Hajnal. *Singularities in monochromatic electromagnetic waves.* PhD thesis, University of Bristol, 1985.

[Haj87a]   J.V. Hajnal. Singularities in the transverse fields of electromagnetic waves. I. Theory. *Proc.R.Soc.Lond.A*, **414**:4433–446, 1987.

[Haj87b]   J.V. Hajnal. Singularities in the transverse fields of of electromagnetic waves. II. Observations on the electric field. *Proc.R.Soc.Lond.A*, **414**:447–468, 1987.

[Hal81]   B.I. Halperin. Statistical mechanics of topological defects. In R. Balian, M. Kléman, and J.-P. Poirier, editors, *Les Houches Session XXV - Physics of Defects*, pages 813–857. North-Holland, Amsterdam, 1981.

[Han96]   J.H. Hannay. Chaotic analytic zero points: exact statistics for those of a random spin state. *J.Phys.A:Math.Gen.*, **29**:L101–L105, 1996.

[Han98a]   J.H. Hannay. Cyclic rotations, contractibility and Gauss-Bonnet. *J.Phys.A:Math.Gen.*, **31**:L321–L324, 1998.

[Han98b]   J.H. Hannay. The Berry phase for spin in the Majorana representation. *J.Phys.A:Math.Gen.*, **31**:L53–L59, 1998.

[Han98c]   J.H. Hannay. The chaotic analytic function. *J.Phys.A:Math.Gen.*, **31**:L755–L761, 1998.

[Han98d]   J.H. Hannay. The Majorana representation of polarization, and the Berry phase of light. *J.Mod.Opt.*, **45**:1001–1008, 1998.

[Hay84]   M. Hayes. Inhomogeneous plane waves. *Arch.Rat.mech.Anal.*, **85**:41–79, 1984.

[HCP75]   J.O. Hirschfelder, A.C. Christoph, and W.E. Palke. Quantum mechanical streamlines. I. Square potential barrier. *J.Chem.Phys.*, **61**:5453–5455, 1975.

[HGB74]   J.O. Hirschfelder, C.J. Goebel, and L.W. Bruch. Quantized vortices around wavefunction nodes. II. *J.Chem.Phys.*, **61**:5456–5459, 1974.

[HGP98]   G. Horvath, J. Gal, and I. Pomozi. Polarization portrait of the Arago point: Video-polarimetric imaging of the neutral points of skylight polarization. *Naturwiss.*, **85**:333–339, 1998.

[HM75]   J.P. Hansen and I.R. McDonald. Statistical Mechanics of dense ionized matter. IV. Density and charge fluctuations in a simple molten salt. *Phys.Rev.A*, **11**:2111–2123, 1975.

[HM86]   J.P. Hansen and I.R. McDonald. *Theory of simple liquids*. Academic Press, 1986.

[HOG89]   E.J. Heller, P. O'Connor, and J. Gehlen. The eigenfunctions of classically chaotic systems. *Phys.Script.*, **40**:354–359, 1989.

[Hol87]   P.R. Holland. Geometry of dislocated de Broglie waves. *Found.Phys*, **17**:345–363, 1987.

[HT76a]   J.O. Hirschfelder and K.T. Tang. Quantum mechanical streamlines. III. Idealized reactive atom-diatomic molecule collision. *J.Chem.Phys.*, **64**:760–785, 1976.

[HT76b]   J.O. Hirschfelder and K.T. Tang. Quantum mechanical streamlines. IV. Collision of two spheres with square potential wells or barriers. *J.Chem.Phys.*, **65**:470–486, 1976.

[Jan87]   B. Jancovici. Charge correlations and sum rules in coulomb systems. i. In F.J. Rogers, editor, *Strongly Coupled Plasma Physics*, pages 349–355. Plenum Publishing Corporation, 1987.

[Kel67]   Lord Kelvin. On vortex atoms. *Phil.Mag.*, **34**:15–24, 1867.

[Kel69]   Lord Kelvin. On vortex motion. *Trans.R.Soc.Ed.*, **25**:217–260, 1869.

[KSMH80]   D. Kramer, H. Stephani, M. MacCallum, and E. Herlt. *Exact solutions of Einstein's field equations*. Cambridge University Press, 1980.

[Lar01]   K.G. Larkin. Natural demodulation of two-dimensional fringe patterns. II. Stationary phase analysis of the spiral phase quadrature transform. *J.Opt.Soc.Am.A*, **18**:1871–1881, 2001.

[LBO01]   K.G. Larkin, D.J. Bone, and M.A. Oldfield. Natural demodulation of two-dimensional fringe patterns. I. General background of the spiral phase quadrature transform. *J.Opt.Soc.Am.A*, **18**:1862–1870, 2001.

[Leb91]   P. Leboeuf. Phase space approach to quantum dynamics. *J.Phys.A:Math.Gen.*, **24**:4575–4586, 1991.

[Lee98]   R.L. Lee. Digital imaging of clear-sky polarization. *App.Opt.*, **37**:1465–1476, 1998.

[LH57a]   M.S. Longuet-Higgins. Statistical properties of an isotropic random surface. *Phil.Trans.R.Soc.A*, **250**:157–174, 1957.

[LH57b]   M.S. Longuet-Higgins. The statistical analysis of a random, moving surface. *Phil.Trans.R.Soc.A*, **249**:321–387, 1957.

[LH58]   M.S. Longuet-Higgins. On the intervals between successive zeros of a random function. *Proc.R.Soc.Lond.A*, **246**:99–118, 1958.

[LH60]   M.S. Longuet-Higgins. Reflection and refraction at a random moving surface. *J.Opt.Soc.Am.*, **50**:838–856, 1960.

[LL75]   L.D. Landau and E.M. Lifshitz. *The classical theory of fields*. Pergamon Press, 1975.

[LL77]   L.D. Landau and E.M. Lifshitz. *Quantum Mechanics (Non-relativistic theory)*. Pergamon Press, 1977.

[Mac27]   T.M. MacRobert. *Spherical Harmonics*. Methuen and Co., London, 1927.

[Maj32]   E. Majorana. Atomi orientati in campo magnetico variabile. *Nuov.Cim*, **9**:43–50, 1932.

[Man67]   L. Mandel. Complex representation of optical fields in coherence theory. *J.Opt.Soc.Am.*, **57**:613–617, 1967.

[MB89]   R.J. Mondragon and M.V. Berry. The quantum phase 2-form near degeneracies: two numerical studies. *Proc.R.Soc.Lond.A*, **424**:263–278, 1989.

[Mec58]   A. Meckler. Majorana formula. *Phys.Rev.*, **111**:1447–1449, 1958.

[Meh89]   M.L. Mehta. *Matrix Theory: Selected topics and useful results*. Les Editions de Physique, Les Ulis, 1989.

[Mer79]   N.D. Mermin. The topological theory of defects in ordered media. *Rev.Mod.Phys.*, **51**:591–648, 1979.

[Mil65]   J.W. Milnor. *Topology from the differentiable viewpoint*. Virginia University Press, 1965.

[Mof69]   H.K. Moffatt. The degree of knottedness of tangled vortex lines. *J.Fluid Mech.*, **35**:117–129, 1969.

[MR92]   H.K. Moffatt and R.L. Ricca. Helicity and the Calugareanu invariant. *Proc.R.Soc.Lond.A*, **439**:411–429, 1992.

[MTW73]   C.W. Misner, K.S. Thorne, and J.A. Wheeler. *Gravitation*. W.H. Freeman and Company, San Francisco, 1973.

[MW95]   L. Mandel and E. Wolf. *Optical coherence and quantum optics*. Cambridge University Press, 1995.

[NB74]   J.F. Nye and M.V. Berry. Dislocations in wave trains. *Proc.R.Soc.Lond.A*, **336**:165–190, 1974.

[Nee97]   T. Needham. *Visual complex analysis*. Oxford University Press, 1997.

[NH87]   J.F. Nye and J.V. Hajnal. The wave structure of monochromatic electromagnetic radiation. *Proc.R.Soc.Lond.A*, **409**:21–36, 1987.

[NHH88]   J.F. Nye, J.V. Hajnal, and J.H. Hannay. Phase saddles and dislocations in two-dimensional waves such as the tides. *Proc.R.Soc.Lond.A*, **417**:7–20, 1988.

[Niv80]   L. Niven. *Convergent Series*. Macdonald Futura, 1980.

[NP62]   E. Newman and R. Penrose. An approach to gravitational radiation by a method of spin coefficients. *J.Math.Phys.*, **3**:566–579, 1962.

[Nye81]   J.F. Nye. The motion and structure of dislocations in wavefronts. *Proc.R.Soc.Lond.A*, **378**:219–239, 1981.

[Nye83a]   J.F. Nye. Lines of circular polarization in electromagnetic wave fields. *Proc.R.Soc.Lond.A*, **389**:279–290, 1983.

[Nye83b]   J.F. Nye. Polarization effects in the diffraction of elecromagnetic waves: the role of disclinations. *Proc.R.Soc.Lond.A*, **387**:105–132, 1983.

[Nye91]   J.F. Nye. Phase gradient and crystal-like geometry in electromagnetic and elastic wavefields. In R.G. Chambers, J.E. Enderby, A. Keller, A. Lang, and J.W. Steeds, editors, *Sir Charles Frank OBE: an eightieth birthday tribute*, pages 220–231. Adam Hilger, Bristol, 1991.

[Nye98]   J.F. Nye. Unfolding of higher-order wave dislocations. *J.Opt.Soc.Am.A*, **15**:1132–1138, 1998.

[Nye99]   J.F. Nye. *Natural focusing and fine structure of light: caustics and wave dislocations.* Institute of Physics Publishing, Bristol, 1999.

[OG83]   E. Ochoa and J.W. Goodman. Statistical properties of ray directions in a monochromatic speckle pattern. *J.Opt.Soc.Am.*, **73**:943–949, 1983.

[OGH87]   P. O'Connor, J. Gehlen, and E.J. Heller. Properties of random superpositions of plane waves. *Phys.Rev.Lett.*, **58**:1296–1299, 1987.

[Pan56]   S. Pancharatnam. Generalized theory of interference and its applications. *Proc.Ind.Acad.Sci.*, **64**:247–262, 1956.

[Par97]   D.A. Park. *The fire within the eye: A historical essay on the nature and meaning of light.* Princeton University Press, 1997.

[Pen60]   R. Penrose. A spinor approach to general relativity. *Ann.Phys.(NY)*, **10**:171–201, 1960.

[Pen79]   R. Penrose. The topology of ridge systems. *Ann.Hum.Gen.*, **42**:435–444, 1979.

[Pen89]   R. Penrose. *The Emperor's New Mind.* Oxford University Press, 1989.

[Pen94]   R. Penrose. *Shadows of the Mind.* Oxford University Press, 1994.

[Pla49]   M. Planck. *Scientific Autobiography and other papers.* Philosophical Library, London, 1949.

[Poi92]   H. Poincaré. *Théorie mathématique de la Lumière*, volume 2. Georges Carré, Paris, 1892.

[PR84a]   R. Penrose and W. Rindler. *Spinors and Space-time, Volume I. Two-spinor calculus and relativistic fields.* Cambridge University Press, 1984.

[PR84b]   R. Penrose and W. Rindler. *Spinors and Space-time, Volume II: Spinor and twistor methods in space-time geometry.* Cambridge University Press, 1984.

[PS78]   T. Poston and I. Stewart. *Catastrophe theory and its applications.* Pitman Publishing Ltd, London, 1978.

[Ray71]   Lord Rayleigh. On the light from the sky, its polarization and colour. *Phil.Mag.*, **41**:107–120, 274–279, 1871.

[Ray89]   Lord Rayleigh. On the character of the complete radiation at a given temperature. *Phil.Mag.*, **27**:460–469, 1889.

[RB85]   M. Robnik and M.V. Berry. Classical billiards in magnetic fields. *J.Phys.A:Math.Gen.*, **18**:1361–1378, 1985.

[Ric44]   S.O. Rice. Mathematical analysis of random noise. *Bell Syst.Tech.J.*, **23**:282–332, 1944.

[Ric45]   S.O. Rice. Mathematical analysis of random noise. *Bell Syst.Tech.J.*, **24**:46–156, 1945.

[Rie70a]   J. Riess. Nodal structure, nodal flux fields, and flux quantization in stationary quantum states. *Phys.Rev.D*, **2**:647–653, 1970.

[Rie70b]   J. Riess. Nodal structure of Schroedinger wave functions and its physical significance. *Ann.Phys.(NY)*, **57**:301–321, 1970.

[Rie87]   J. Riess. Quantised vortex motion through rings in quantum mechanics. *J.Phys.A:Math.Gen.*, **20**:5179–5188, 1987.

[RSB99]   R.L. Ricca, D.C. Samuels, and C.F. Barenghi. Evolution of vortex knots. *J.Fluid Mech.*, **391**:29–44, 1999.

[Sak94]   J.J. Sakurai. *Modern Quantum Mechanics (Revised Edition).* Addison-Wesley, 1994.

[SBS01]   A.I. Saichev, K.-F. Berggren, and A.F. Sadreev. Distribution of nearest distances between nodal points for the Berry function in two dimensions. *Phys.Rev.E*, **64**(036222), 2001.

[Sch80]   B.F. Schutz. *Geometrical methods of mathematical physics.* Cambridge University Press, 1980.

[Sch85]   B.F. Schutz. *A first course in general relativity.* Cambridge University Press, 1985.

[SDAP97]   N.B. Simpson, K. Dholakia, L. Allen, and M.J. Padgett. Mechanical equivalent of spin and orbital angular momentum: an optical spanner. *Opt.Lett.*, **22**:52–54, 1997.

[Sei80]   H. Seifert. Topology of 3-dimensional fibred spaces. In H. Seifert and W. Threllfall, editors, *A textbook of topology.* Academic Press, 1980.

[SF94a]   N. Shvartsman and I. Freund. Vortices in random wave fields: Nearest neighbor anticorrelations. *Phys.Rev.Lett.*, **72**:1008–1011, 1994.

[SF94b]   N. Shvartsman and I. Freund. Wavefield phase singularities: near-neighbour correlations and anticorrelations. *J.Opt.Soc.Am.A*, **11**:2710–2718, 1994.

[SF95]   N. Shvartsman and I. Freund. Speckle spots ride phase saddles sidesaddle. *Opt.Commun.*, **117**:228–234, 1995.

[SHH77]   P.A.G. Scheuer, J.H. Hannay, and P.J. Hargrave. A note on the interpretation of polarization maps. *Mon.Not.R.astr.Soc.*, **180**:163–167, 1977.

[SL68a]   F.H. Stillinger and R. Lovett. General restriction on the distribution of ions in electrolytes. *J.Chem.Phys.*, **49**:1991–1994, 1968.

[SL68b]   F.H. Stillinger and R. Lovett. Ion-pair theory of concentrated electrolytes. I. Basic concepts. *J.Chem.Phys.*, **48**:3858–3868, 1968.

[Sos98]   M.S. Soskin, editor. *Proceedings of International Conference on Singular Optics*, volume 3487. SPIE, 1998.

[SS96]   Y.Y. Schechner and J. Shamir. Parameterization and orbital angular momentum of anisotropic dislocations. *J.Opt.Soc.Am.A*, **13**:967–973, 1996.

[ST80]   H. Seifert and W. Threlfall. *A textbook of topology.* Academic Press, 1980.

[SV01]   M.S. Soskin and M.V. Vasnetsov, editors. *Singular Optics (Optical Vortices): Fundamentals and Applications*, volume 4403. SPIE, 2001.

[SW64]   R.F. Streater and A.S. Wightman. *PCT, spin and statistics, and all that.* W A Benjamin, New York, 1964.

[SW89]   A. Shapere and F. Wilczek, editors. *Geometric phases in physics.* World Scientific, 1989.

[Syn58]   J.L. Synge. *Relativity: The special theory.* North-Holland, Amsterdam, 1958.

[Syn64]   J.L. Synge. The Petrov classification of gravitational fields. *Comm.Dubl.Inst.Adv.St.*, **15**, 1964.

[Tit48]   E.C. Titchmarsh. *Introduction to the theory of Fourier integrals.* Oxford University Press, 2nd edition, 1948.

[TT90]   D.R. Tilley and J. Tilley. *Superconductivity and superfluidity.* Adam Hilger, Bristol, 3rd edition, 1990.

[vdH49]   H.C. van de Hulst. Scattering in the atmospheres of the earth and planets. In G.P. Kuiper, editor, *The atmospheres of the Earth and planets*, pages 49–111. University of Chicago Press, 1949.

[Ver92]   J. Verne. *Around the world in 80 days.* Penguin, 1992.

[VMK88]   D.A. Varshalovich, A.N. Moskalev, and V.K. Khersonskii. *Quantum Theory of Angular Momentum.* World Scientific, Singapore, 1988.

[VS99]   M. Vasnetsov and K. Staliunas, editors. *Optical vortices.* Nova Science, Commack, NY, 1999.

[Wel33]   H.G. Wells. *The scientific romances of H.G. Wells.* Victor Gollancz Ltd, London, 1933.

[WH82]   A. Weinrib and B.I. Halperin. Distribution of maxima, minima, and saddle points of the intensity of laser speckle patterns. *Phys.Rev.B*, **26**:1362–1368, 1982.

[Whe33]   W. Whewell. Essay towards a first approximation to a map of cotidal lines. *Phil.Trans.R.Soc.*, **123**:147–236, 1833.

[Whe36]   W. Whewell. Researches on the tides. Sixth Series. On the results of an extensive system of tide observations on the coasts of Europe and America in June 1835. *Phil.Trans.R.Soc.*, **126**:289–341, 1836.

[Whi11]   W. Whitman. *Leaves of Grass*. D. Appleton and Co., London and New York, 1911.

[Wig39]   E.P. Wigner. On unitary representations of the inhomogeneous Lorentz group. *Ann.Math.*, **40**:149–204, 1939.

[Wig59]   E.P. Wigner. *Group theory and its application to the quantum mechanics of atomic spectra*. Academic Press, 1959.

[Win80]   A.T. Winfree. *The geometry of biological time*. Springer, 1980.

[Win87]   A.T. Winfree. *When time breaks down*. Princeton University Press, 1987.

[Win95]   A.T. Winfree. Persistent tangles of vortex rings in excitable media. *Physica D*, **84**:126–147, 1995.

[Woa00]   G. Woan. *The Cambridge handbook of physics formulas*. Cambridge University Press, 2000.

[Wol99]   S. Wolfram. *The Mathematica book*. Cambridge University Press, 1999.

[Wri77]   F.J. Wright. *Wavefield singularities*. PhD thesis, University of Bristol, 1977.

[Wri79]   F.J. Wright. Wavefront dislocations and their analysis using catastrophe theory. In W. Güttinger and H. Eikemeier, editors, *Structural Stability in Physics*, pages 141–156. Springer-Verlag, 1979.

[WS83a]   A.T. Winfree and S.H. Strogatz. Singular filaments organize chemical waves in three dimensions. I. Geometrically simple waves. *Physica D*, **8**:35–49, 1983.

[WS83b]   A.T. Winfree and S.H. Strogatz. Singular filaments organize chemical waves in three dimensions. II. Twisted waves. *Physica D*, **9**:65–80, 1983.

[WS83c]   A.T. Winfree and S.H. Strogatz. Singular filaments organize chemical waves in three dimensions. III. Knotted waves. *Physica D*, **9**:333–345, 1983.

[WS84]   A.T. Winfree and S.H. Strogatz. Singular filaments organize chemical waves in three dimensions. IV. Wave taxonomy. *Physica D*, **13**:221–233, 1984.

[WWS85]   A.T. Winfree, E.M. Winfree, and H. Seifert. Organizing centers in a cellular excitable medium. *Physica D*, **17**:109–115, 1985.

[WY75]   T.T. Wu and C.N. Yang. Concept of nonintegrable phase factors and global formation of gauge fields. *Phys.Rev.D*, **12**:3845–3857, 1975.

[You02]   T. Young. On the theory of light and colours (The Bakerian lecture). *Phil.Trans.R.Soc.*, **20**:12–48, 1802.