3. Demodulation of Two-Dimensional Electromagnetic Intensity

The farther backward you can look,

the farther forward you are likely to see.

Winston Churchill

3.1 Introduction

Following on from the one-dimensional, interferometric, envelope and phase estimation in chapter 2, a problem in two-dimensional wavefield analysis arose in experiments in microscopic light sectioning using planar waveguides. A number of researchers have considered the non-interferometric reconstruction of complex wavefields from intensity measurements.1,2 Previously it has been shown that for partially coherent systems this is not, in general, possible because different wavefields can exhibit identical intensity distributions.3-5 The more restricted problem of finding the complex wavefield corresponding to the three-dimensional intensity in a coherent system may be soluble by phase retrieval techniques but is not directly soluble unless the phase variation is suitably constrained: essentially the transverse phase derivative must be smoothly varying.6,7
A particular case of the general problem that is demonstrably soluble by a direct method is considered. The particular case is essentially an optical wavefield with cylindrical symmetry. Such symmetry reduces the problem from three to two dimensions, resulting in a well-posed inverse problem. The resulting two-dimensional problem may be considered as a demodulation problem or as a typical phase retrieval problem with a very specific support constraint due to the wave equation. Phase retrieval in two-dimensions results in a unique solution in most practical cases. However, the calculation invariably relies on an iterative error-reducing algorithm originally proposed by Gerschberg and Saxton. More recently, the convergence of the algorithm has been accelerated, but the method remains very sensitive to noise in the starting data.

The importance of support constraints in the avoidance of ambiguity is well known. Using the special support constraints of the wave equation a direct method of phase retrieval from the intensity of two-dimensional coherent wavefield can be developed. The solution presented is not merely a theoretical curiosity; systems with the required symmetry occur in slab waveguides and slit illumination systems.

It is, perhaps, surprising to note that a significant body of research – broadly contained under the description “diffraction tomography” – has covered much of the background to the theory presented in this chapter, yet has not proposed such a method before (as far as can be ascertained from the available literature). So, for example, the circular arc support constraint is fundamental to the Fourier representation of optical tomographic sampling. However, most tomographic or inverse optical methods explicitly require measurement or estimation of the coherent field as input data for the back-projection or back-propagation reconstruction
algorithms.\textsuperscript{15, 17-19} Therefore, in practice, a preliminary step of phase retrieval from intensity measurements is required. A number of phase retrieval methods have been proposed such as interferometric (or holographic) measurement,\textsuperscript{20-22} iterative retrieval from single intensity planes, and direct retrieval from two intensity planes\textsuperscript{19, 21} (and more recently of interest to adaptive optics and x-ray phase imaging). It may very well be that the number of sampled intensity points required by our proposed method was previously considered so extreme that any such solution would be considered impractical. The comparable methods in two dimensions only require one or two lines of sampled intensities, whereas a sequence of lines covering a two dimensional patch is required for the proposed method. The collection of such a 2D dataset is not considered here, but it is not inconceivable that a carefully designed optical imaging system could collect all the required intensity data simultaneously. Not all tomographic systems require coherent field data: for example the 3D reconstruction of fluorescent stained objects is essentially an incoherent process,\textsuperscript{23-25} although objects with varying refractive index do require phase information associated with the optical field.\textsuperscript{26}

In the following sections the Fourier representation of the problem that leads to a convenient coordinate transformation which converts the problem into estimation of a 2D separable function is outlined. Computational aspects are considered and an experimental (computer simulation) verification appear in the later sections before some conclusions are drawn.
3.2 Fourier Representation of Problem

Access to the field intensity at all points in the $x$-$z$ plane is assumed initially. Details of sampling and the limited extent of measurements are ignored in this preliminary investigation. The assumption of cylindrical symmetry results in a field independent of the $y$ coordinate. A scalar, quasi-monochromatic optical field $E$ propagating in the $z$ direction with field intensity $I$ given by

\[ I(x, y, z) \equiv |E(x, y, z)|^2 = |f(x, y, z)|^2 = I(x, z) = |f(x, z)|^2 \tag{3.1} \]

The notation for the intensity function $|f(x, y, z)|^2 = I(x, y, z)$ is used here for convenience. The field considered sufficiently far from apertures or other obstructions for near-field contributions to be negligible. In such a case the Debye approximation is appropriate if the Fresnel number of the system is very large. The propagation medium is a linear, uniform and isotropic with refractive index 1. Under such conditions, the field function $f(x, y, z)$ satisfies a simple wave equation, more specifically the Helmholtz equation:

\[ \nabla^2 f(x, y, z) + k^2 f(x, y, z) = 0. \tag{3.2} \]

The wave number and wavelength are related as usual by $k\lambda = 2\pi$. It is worth noting that the approximations used are not strictly valid for high aperture optical systems unless the vectorial components are explicitly included. However, the results are
valid for high aperture scalar field. The Fourier transform (FT) of the field function (sometimes known as the angular spectrum) defined for the spatial frequency coordinates \((m,n,s)\) is given by

\[
F(m,n,s) = \iiint f(x,y,z) \exp[-2\pi i(mx + ny + sz)] \, dx \, dy \, dz.
\] (3.3)

Fourier transforming equation \([3.2]\) reveals that the transform of the field only exists on the surface of the Ewald sphere.\(^{29}\)

\[
F(m,n,s) \left[ k^2 - (2\pi)^2 (m^2 + n^2 + s^2) \right] = 0.
\] (3.4)

The function \(F(m,n,s)\) is sometimes referred to as the 3D-pupil function, or more generally as the angular spectrum. The lack of variation of \(f(x,y,z)\) in the \(y\) direction, i.e. the cylindrical symmetry ensures the further restriction of \(F\) to an arc of a circle, radius \(1/\lambda\), in the \((m,s)\) plane

\[
F(m,n,s) \left[ 1 - \lambda^2 \left( m^2 + n^2 + s^2 \right) \right] \delta(s) = 0 = F(m,s) \left[ 1 - \lambda^2 \left( m^2 + s^2 \right) \right].
\] (3.5)

The measurements of the field intensity (modulus squared of the field) yield an FT that is the autocorrelation of \(F(m,s)\). If \(F(m,s)\) exists only on a circular arc, then its
autocorrelation exists only on a bow-shaped (or petal-shaped) region bounded by similar circular arcs, as shown in figure 3.1 (b) (see also Figure 3.4).

Figure 3.1(a)
Region of support for the pupil function $F(m,s)$.

Figure 3.1(b)
Regions of support for the autocorrelation of the pupil function.

In this case, the arc (figure 3.1 (a)) corresponds to the angular spectrum of an optical wave propagating from left to right.
3.3 Coordinate Transform

When considering the autocorrelation of a function with support on a circular arc it is more convenient to use a coordinate measuring the angular position along the arc. Figure 3.2 shows the angular parameter defined by the following coordinate transform:

\[
\begin{align*}
    m\lambda &= \sin \theta_1 - \sin \theta_2 \\
    s\lambda &= \cos \theta_2 - \cos \theta_1
\end{align*}
\]

(3.6)

This coordinate transform is essentially the same as that adopted by Sheppard to describe microscope imaging of 3D objects in transmission.

In equation (3.3) the Fourier transform of the wavefield \( f(x, z) \) is the 2D angular spectrum \( F(m, s) \) and is directly analogous to the system pupil function. By transforming to the new coordinate system of equation (3.6) the pupil function \( \tilde{F}(\theta) \)
parameterised by the angle $\theta$ is obtained. Similarly, the transform of the 2D-intensity pattern $g(x, z) = I(x, z) = |f(x, z)|^2$ is given by

$$G(m, s) = \iint g(x, z) \exp[-2\pii(mx + sz)]dxdz$$

$$= \iint F(m', s')F^*(m' - m, s' - s)dm'ds'$$

(3.7)

The above function corresponds to the 3D OTF (optical transfer function) first proposed for the analysis of optical imaging systems by Frieden.$^{31}$ Interestingly, Mertz considered the problem for cylindrical imaging in 1965.$^{32}$ Again, the resulting function can be transformed using equation (3.6) to give $\tilde{G}(\theta_1, \theta_2)$. The omission of the wavelength parameter $\lambda$ in the following analysis is equivalent to having a priori knowledge of the exact value. Figure 3.2 represents the overlap of two arcs as they are correlated. Figure 3.3 shows more detail for the general case of arcs with finite thickness (i.e. non-zero bandwidth). The overlap area is related to the angular difference $2\theta_1 - 2\theta_2$. Treating the arc-supported functions as generalised functions then gives the simple overlap integral

$$\tilde{G}(\theta_1, \theta_2) = \frac{\tilde{F}(\theta_1)\tilde{F}^*(\theta_2)}{\sin(\theta_1 - \theta_2)}.$$  

(3.8)

Counter-propagating waves have been ignored, which would otherwise give an additional overlap term at $(\theta_1 + \pi)$, and $(\theta_2 + \pi)$.  


Figure 3.3
The rhombic overlap region in the autocorrelation of $F(m,s)$.

Clearly $\tilde{G}(\theta_1, \theta_2)$ is a separable function of $\tilde{F}(\theta_1)$ and $\tilde{F}(\theta_2)$ which occupies a square region $-\Theta < \theta_1 < \Theta$ and $-\Theta < \theta_2 < \Theta$, where the optical system semi-aperture is defined by the angle $\Theta \leq \pi/2$. Figure 3.4 shows the coordinate transformation $(s,m) \rightarrow (\theta_1, \theta_2)$.

A similar analysis by Sheppard et al. derives a square region with an upper triangular portion related to a convolution rather than a correlation. Equation (3.8) is effectively the recipe for deconvolution of $G(m,s)$. The constraints of the wave equation have ensured that the autocorrelation only involves the overlap of two points (located at $\theta_1$ and $\theta_2$) which allows subsequent separation of the simple product $\tilde{F}(\theta_1)\tilde{F}^r(\theta_2)$.

Equation (3.8) contains a singularity at $\theta_1 = \theta_2$. The singularity only occurs in the limit of monochromaticity.
In practice, the near-singularity affects the noise sensitivity of phase calculations in the region $\theta_1 = \theta_2$, as shown in the next section.

![Diagram of coordinate transformation](image)

Figure 3.4  
Effect of coordinate transformation for a typical grid pattern.

3.4 Computational Procedure

The preceding sections have outlined an intuitively inspired method of undoing the autocorrelation associated with the modulus squaring operation. The actual computation scheme developed to numerically model the process is now outlined. In the continuous domain the combination of equations (3.6) and (3.7) gives:

$$G(m, s) = \tilde{G}(\theta_1, \theta_2) = \int g(x, z) \exp\left[-\frac{2\pi i}{\lambda}\left([\sin \theta_1 - \sin \theta_2]x + [\cos \theta_2 - \cos \theta_1]z\right)\right] dx dz$$

(3.9)
From this the underlying angular spectrum $\tilde{F}(\theta)$ is available in a 2D separable form:

$$\tilde{F}(\theta_1)\tilde{F}^*(\theta_2) = \sin(\theta_1 - \theta_2)|\tilde{G}(\theta_1, \theta_2)|. \quad (3.10)$$

The main problem in the computation is the interpolation intrinsic to the coordinate transform of equation [3.6]. The coordinate transformation occurs after the initial step of Fourier transforming the measured intensity distribution. Typically, the Fourier transformation is performed as a discrete Fast Fourier transform (FFT) with all the speed advantages that implies. However, by utilising a discrete Fourier transformation with non-uniformly spaced indices both the Fourier and the coordinate transformations can be combined into one. The redundancy properties needed for the FFT decomposition are lost, but perfect sinc-function based (bandlimited) interpolation results for all the new coordinate indices. The discrete Fourier transform (DFT) for estimating the angular spectrum of $g(x, z)$ is as follows:

$$G_{\rho, \sigma} = G(\rho/[N\Delta], \sigma/[N\Delta]) = \frac{\Delta^2}{N} \sum_{q=0}^{N-1} \sum_{p=0}^{N-1} g(p\Delta, q\Delta) \exp(2\pi[pp + q\sigma]/N) \quad (3.11)$$

where $g_{p,q} = g(p\Delta, q\Delta)$, and both p and q are integers representing the intensity g on a square grid of sample points separated by $\Delta$. Typical examples of $g(x, z)$ and $|G(m, s)|$ are shown in Figure 3.5 and figure 3.6. Both figure 3.5 and figure 3.6 represent the functions as grey levels in an image. The grey level scales are nonlinear.
to enhance the low values and make salient detail visible. Equation (3.11) represents the function $G$ on a square grid with integers indices $\rho$ and $\sigma$.

Figure 3.5
Greyscale plot of a typical intensity pattern $g(x,z)$. The quartic root of the intensity is displayed to enhance low level features of interest. This particular pattern corresponds to the focal region of an aberrated cylindrical wave.

Figure 3.6
Greyscale plot of the modulus of the Fourier transformed intensity, $|G(m,s)|$. The greyscale is nonlinear to enhance low-level features of interest. The predicted correlation “petal” outline is clearly shown. Note that $G(m,s)$ is, in general, complex.
Combining the coordinate transform, so that the transform is evaluated at inter-sample points gives the following:

\[
\tilde{F}_\alpha \tilde{F}_\beta^* = \tilde{G}(\alpha, \beta) = \frac{\Delta^2}{N} \sin\left(\frac{2\pi(\alpha - \beta)}{N}\right)
\]

\[
\times \sum_{p=\frac{N}{2}+1}^{N-1} \sum_{q=\frac{N}{2}+1}^{N-1} g_{p,q} \exp\left\{\frac{2\pi L}{N} \left[ p \left( \cos\left(\frac{2\pi \alpha}{N}\right) - \cos\left(\frac{2\pi \beta}{N}\right) \right) + q \left( \sin\left(\frac{2\pi \beta}{N}\right) - \sin\left(\frac{2\pi \alpha}{N}\right) \right) \right]\right\}
\]

(3.12)

The discrete form of the coordinate transform has the following scaling:

\[
\begin{align*}
\theta_1 &= \frac{2\pi \alpha}{N} \\
\theta_2 &= \frac{2\pi \beta}{N} \\
\lambda &= \frac{L}{N\Delta}
\end{align*}
\]

(3.13)

which gives N-1 samples over the range \(-N/2 + 1 \leq \alpha < N/2 - 1\), where \(\alpha\) is the integer. A similar relation applies to \(\beta\), so that a mapping into a 2D grid with angular coordinates symmetrically spanning \(2\pi\) in each direction results. Once \(\tilde{G}(\theta_1, \theta_2)\) is calculated variety of simple statistical techniques can be applied to extract \(\tilde{F}(\theta)\) depending on the model of expected noise in the system. So for example (in the noise-free case) the function \(\tilde{G}(\theta_1, \theta_2)\) can be normalised, removing the singularity in the process, and then integrating over one coordinate:

\[
\int_{-\theta}^{\theta} \tilde{G}(\theta_1, \theta_2) |\sin(\theta_1 - \theta_2)| d\theta_1 = \tilde{F}(\theta) \int_{-\theta}^{\theta} \tilde{F}(\theta_2) d\theta_2.
\]

(3.14)
Hence, both the amplitude and phase (i.e. the complex amplitude) of $\mathcal{F}(\theta)$ can be found for a coherent field. Alternatively, methods (such as correlation or matched filtering) which exploit the large degree of redundancy in the 2D separable function can be utilised to good effect. In systems with partial coherence a unique phase cannot be associated with each part of the angular spectrum but some overall statistical parameters may be calculated in a similar manner.

### 3.5 Experimental Verification

**3.5.1 A preliminary verification of the method was implemented as follows:**

I. A 2D-pupil function (angular spectrum) with constant magnitude and known phase variation was generated; see figure 3.7.

II. The corresponding 2D-field pattern was calculated using the FFT in the Debye approximation.

III. The corresponding modulus-squared (intensity) pattern was calculated; see figure 3.5.

IV. The direct phase retrieval algorithm embodied in equation \[3.12\] was applied to the 2D-intensity pattern resulting in the magnitude and phase patterns shown in figure 3.8.

V. Integration along the $\theta_2$ direction was used to estimate the phase component of $\mathcal{F}(\theta_1)$ according to equation \[3.14\]

VI. Steps III to V were repeated for the same intensity distribution with uniform random noise added.
Figure 3.7
Greyscale plot of the initial angular spectrum $F(m,s)$ as a magnitude (left) and a phase (right) component. The greyscales are linear in this case. Note that the circular arc subtends $\pi$ radians and has been given a narrow Gaussian profile to reduce line aliasing artefacts. The imposed phase modulation has a sinusoidal form with five periods over the arc length. The linear greyscale is shown far left.

Figure 3.8
Greyscale plot of $F(\theta_1)F^*(\theta_2)$ from the direct phase retrieval algorithm of equation (12). The magnitude is on the left and the phase on the right. Note that the magnitude is near constant (light grey) over the upper right quadrant corresponding to the range $-\pi/2 < \theta_1 < \pi/2$ in this implementation. The phase component clearly shows the five period grid variation expected in this example. The patterns have Hermitian symmetry and could be adequately defined over a region half this area.
3.5.2 Simulation Details

The angular spectrum used (figure 3.7) has a radius of 32 pixels in a sampled region of 128x128 pixels. The circular arc has a narrow Gaussian profile to reduce line aliasing artifacts. The spectrum has an imposed phase variation of sinusoidal form with five periods over the semicircular arc. Note that the five period phase function cannot be represented by a low order polynomial, and may be expected to defeat phase retrieval schemes that inherently require slow phase variation to succeed. The sampling is approximately 4 pixels per fringe, giving 32 fringes over the 128-pixel frame. The resulting intensity distribution just satisfies the Nyquist sampling criterion. Dynamic range and measurement quantisation are also important, but such issues are avoided in this example with the use of floating point numbers.

The direct phase retrieval algorithm requires of order $N^4$ operations to compute for a $N \times N$ image. In this case $N = 128$ and approximately $10^9$ floating point operations are required. More sophisticated, statistically based methods can be used to extract the 2D separable function given by equation (3.12). However, the simplistic analysis of equation (3.14) gives results as shown in figure 3.9.

In fact, there is some error in the range of the recovered phase: the calculation gives a $\pm 0.43$ radian range, whereas the simulation actually used a range of $\pm 0.50$ radian, implying a degree of attenuation in the reconstruction. Because the 1-D information is uniformly spread over 2-D as a separable function, there is a large degree of degeneracy in the output, which could be utilised to reduce noise effects. Alternatively, only a small section of the output shown in figure 3.8 needs to be calculated to estimate the phase when noise is not a concern. Values along the
diagonal of figure 3.8 corresponding to $\alpha = \beta$ have a zero weight from equation \[ (3.12) \] and do not contribute useful amplitude or phase information.

Uniform random noise with a range equivalent to $\pm 2\%$ of the peak intensity value was added to the original intensity distribution and the direct phase retrieval method repeated. This is a significant amount of noise compared to the likely quantization noise, for example. Figure 3.10 shows the noise dominating the intensity distribution in all regions more than a few wavelengths from the peak intensity. After applying equation \[ (3.12) \] a separable function shown in Figure 3.11 is obtained. The noise effects appear to be well spread and the correct underlying phase is still visible. Application of the simple projection of equation \[ (3.14) \] gives the pupil phase function in Figure 3.12. The original pupil phase is still close to its correct form with some attenuation as before. More simulations are needed to better quantify this effect.

![Figure 3.9](image)

The phase profile recovered using equation 3.14. The predicted five period sinusoid is clearly recovered. The 64 samples cover the angular coordinate range $-\pi/2 < \theta < \pi/2$. 
Figure 3.10
Greyscale plot of an intensity pattern with a uniform random noise. The random noise range is ±2% of the peak intensity. The quartic root of the intensity is displayed to enhance low-level features of interest.

Figure 3.11
Greyscale plot of $\tilde{F}(\theta_1)\tilde{F}(\theta_2)$ from the direct phase retrieval algorithm applied to the noisy intensity map. The magnitude (left) and phase (right) areas shown correspond to the top right quadrants in figure 8.
Figure 3.12
The recovered phase profile from the noisy intensity map. Again the predicted five period sinusoid is clear. The 64 samples cover the angular coordinate range $-\pi/2 < \theta < \pi/2$.

3.6 Conclusion

A direct method for calculating the phase of a wavefield in a cylindrically symmetric optical system has been presented, and the initial simulation results are in close agreement with predictions. The method relies on the Fourier transform of the wavefield intensity, which is then remapped into a new coordinate system. The combination of Fourier transform and remapping conveniently overcomes interpolation difficulties but results in a significant computational load. Much work remains to fully characterise the method and explore its limitations, especially in the presence of noise, although initial simulations indicate a significant insensitivity to additive intensity noise. From the point of view of phase retrieval in 2-D there are some unusual support constraints with physical significance, and a comparison with iterative retrieval schemes may prove illuminating. The method is applicable to certain optical systems, such as the light emitted from slab waveguides and other
anamorphic systems. Measurements required to characterise such systems could be obtained quickly by the use of a 2-D photodetector array in a suitably arranged imaging system. The method presented is an interesting alternative to conventional interferometric phase measurement methods.

3.6.1 Limitations of this chapter

The next logical step in this work would be to re-synthesize the wavefield complex amplitude from the estimated pupil function. The squared modulus of the wavefield could then be compared with the original measured intensity. This final step has not been performed here. Instead, the comparisons have been based on the initial pupil phase function.

3.7 Connections

3.7.1 Phase Space Representations

Soon after the submission of the work in this chapter the authors (Larkin and Sheppard) realised that the quantity defined in equation (3.8) was a high numerical aperture version of the spectral correlation function. Papoulis refers to this function in his work on the ambiguity function in optics. Subsequently the connections have been investigated in two papers.

Wolf, Alonso, and Forbes have developed the theory of Wigner functions of Helmholtz wavefields. The Wigner distribution, the spectral correlation function, and the ambiguity function are simply related to each other by multidimensional Fourier transforms. Further inter-relations involving projections (or marginal
distributions, as they are known) can be used to reduce the dimensionality of the phase space representation.

### 3.7.2 Symmetry and Anisotropy

Although the method presented assumes $\Theta \leq \pi/2$, it has no preferred orientation, unlike certain 2D phase retrieval methods that implicitly assume some rectangular symmetry in the signal bandlimits. If the wave propagation is primarily in a direction other than from left to right, then the regions in the spectral correlation function shift appropriately. Because the high-angle spectral correlation function is defined on a circle (in 2-D) the angular shift is cyclic. Figure 3.8 showed the spectral correlation function for a left-right propagating field. In figure 3.13(a) the pupil function for an up-down propagating field is shown. To make the wavefield distribution in figure 3.13(b) more interesting and widely dispersed, it is generated from a highly variable pupil phase function. Figure 3.13(c) shows the result of applying the direct phase retrieval method to the wavefield intensity. The distribution has shifted by $\pi/2$ in both $\theta_1$ and $\theta_2$ axes. Figure 3.13(d) shows the effect of using the dimensional degeneracy of the spectral correlation function to reduce the underlying uncertainty in the complex function estimate (plotted in pseudocolour to enhance the regular features)
Figure 3.13(a), pupil function magnitude and phase.

Figure 3.13(b), wavefield magnitude and phase.

Figure 3.13(c), spectral correlation function, magnitude and phase.

Figure 3.13(d), spectral correlation function, statistically enhanced pseudocolour representation of magnitude and phase.
3.8 Acknowledgement

I would like to thank Michael Oldfield for writing the software to evaluate equation (3.12).

3.9 References and notes


21 M. H. Maleki, and A. J. Devaney, “Phase retrieval in inverse scattering,”
SPIE Annual Symposium, San Diego, (1992),

22 M. H. Maleki, and A. J. Devaney, “Holographic techniques for inverse
scattering and tomographic imaging,” IS&T/SPIE Symposium on Electronic
Imaging Science and Technology, San Jose, (1994),


24 P. J. Shaw, D. A. Agard, Y. Hiraoka, and J. W. Sedat, “Tilted view
reconstruction in optical microscopy”, Biophysical Journal 55, 101-110,
(1989).

microtomography: multi-angle image acquisition and 3D digital
reconstruction,” Three-Dimensional Microscopy:Image Acquisition and
Processing III, IS&T/SPIE Symposium on Electronic Imaging Science and

26 T. C. Wedberg, and J. Stamnes, “Experimental examination of the
quantitative imaging properties of optical diffraction tomography.”, J. Opt.
In practice we only have access to discretely sampled measurements over a finite region of space. It is a straightforward procedure to include these limitations and the computation artefacts that follow. The main effect, due to the finite area is a blurring in the angular spectrum domain.


