Phase retrieval from the magnitude of the Fourier transforms of nonperiodic objects

J. Miao, D. Sayre,* and H. N. Chapman†

Department of Physics, State University of New York at Stony Brook, Stony Brook, New York 11794

Received October 2, 1997; revised manuscript received January 15, 1998; accepted January 21, 1998

It is suggested that, given the magnitude of Fourier transforms sampled at the Bragg density, the phase problem is underdetermined by a factor of 2 for 1D, 2D, and 3D objects. It is therefore unnecessary to oversample the magnitude of Fourier transforms by $2 \times$ in each dimension (i.e., oversampling by $4 \times$ for 2D and $8 \times$ for 3D) in retrieving the phase of 2D and 3D objects. Our computer phasing experiments accurately retrieved the phase from the magnitude of the Fourier transforms of 2D and 3D complex-valued objects by using positivity constraints on the imaginary part of the objects and loose supports, with the oversampling factor much less than 4 for 2D and 8 for 3D objects. Under the same conditions we also obtained reasonably good reconstructions of 2D and 3D complex-valued objects from the magnitude of their Fourier transforms with added noise and a central stop. © 1998 Optical Society of America [S0740-3232(98)00406-2]

OCIS codes: 100.5070, 100.0100, 070.2590.

1. INTRODUCTION

In a number of fields, such as x-ray crystallography, x-ray diffraction, electron diffraction, neutron diffraction, astronomy, and remote sensing, only the magnitude of the Fourier transform can be measured and the phase of the Fourier transform is lost, which raises the well-known phase problem. In this paper we focus mainly on the phase problem of nonperiodic objects. The first widely accepted phase algorithm mainly for nonperiodic objects was put forward in 1971 by Gerchberg and Saxton.1 The idea is that, if partial information about the magnitude of the object density as well as about the magnitude of the object’s Fourier transform is supplied, the phase information may be recovered. In 1978 Fienup developed two phasing algorithms based partially on the Gerchberg–Saxton algorithm: the error-reduction iterative algorithm and the input–output iterative algorithm.2 Fienup modified the Gerchberg–Saxton algorithm by using finite supports and positivity constraints in real space instead of the magnitude of the object density to retrieve the phase of real and positive objects. In subsequent years, Fienup’s algorithms have been successfully implemented in various fields to determine the phase function of real and positive objects from the magnitude of their Fourier transforms.3–9

The uniqueness of the phase of multidimensional (>2D) real and positive objects with finite supports was shown theoretically in 1979 by Bruck and Sodin10 and in 1982 by Hayes.11 Following some partial ideas proposed in 1947 by Boyes-Watson et al.12 and in 1952 by Sayre,13 in 1982 Bates developed the oversampling method to retrieve the phase from the magnitude of the Fourier transform,14 and it turned out that Fienup’s algorithms could be regarded as examples of that method. Using the argument that the autocorrelation function of any sort of image is twice the size of the image itself in each dimension, Bates stated that the phase information could be recovered by oversampling the magnitude of a Fourier transform that is twice as fine as the Bragg density, i.e., 2× oversampling in each dimension: 4× for two dimensions and 8× for three dimensions.14,15 In 1996 Millane and Stroud16 showed that a 4× oversampling of the Fourier magnitude is sufficient to uniquely determine a multidimensional (>2D) image.16 As will be seen, we, along with Szőke,17 actually go a little further in the same direction. Regarding notation, we define as Bragg density the density of Bragg peaks that would be produced if the nonperiodic structure were turned into a crystal by repetition of the structure with contact but without overlap.

In 1984 Barakat and New sam showed that for complex-valued objects the nonuniqueness of the phase is logically rare.18 Thereupon Bates and Tan in 1985 (Ref. 19) and Lane in 1987 (Ref. 20) concluded that, because of the loss of the positivity constraints, the phase retrieval from the magnitude of the Fourier transform of complex-valued objects is much more difficult than that from real and positive objects. In 1987 Fienup demonstrated the possibility of reconstructing some special complex-valued objects by using their tight-support constraints, where tight support means the true boundary of an original object.21

In this paper we present our understanding of the phase problem as carried out in our work on the possible extension of x-ray crystallography to noncrystals.7,22 In Section 2 we suggest that, given the magnitude of a Fourier transform sampled at the Bragg density, the phase problem is underdetermined by a factor of 2 for one-dimensional (1D), two-dimensional (2D), and three-dimensional (3D) objects. Thus, at least in principle, oversampling the magnitude of a Fourier transform by $2 \times$ in each dimension is unnecessary for retrieving the phase of 2D and 3D objects. In Section 3 we propose that the positivity constraints on the imaginary part of complex-valued objects can be used to retrieve the phase from the magnitude of their Fourier transforms, which makes it much easier to reconstruct complex-valued objects. In Section 4 we present a few successful computer phasing experiments to retrieve the phase of 2D and 3D...
complex-valued objects by oversampling the magnitude of their Fourier transforms that are less than \(2 \times \) in each dimension and using positivity constraints on the imaginary part. In addition, we study the phase retrieval of 2D and 3D complex-valued objects with noise and a central stop.

2. UNIQUENESS OF THE PHASE PROBLEM

Given the density of an object \( f(x) \), its Fourier transform \( F(k) \) is given by

\[
F(k) = \int_{-\infty}^{\infty} f(x) \exp(2\pi i k \cdot x) dx,
\]

(1)

where \( x = (x_1, x_2, x_3) \), i.e., spatial coordinates in image space, and \( k = (k_1, k_2, k_3) \), i.e., spatial-frequency coordinates in Fourier space. In practice, we discretize the object and the Fourier transform by arrays. By using the conventional sampling, we rewrite Eq. (1) and get

\[
F(k) = \sum_{x=0}^{N-1} f(x) \exp(2\pi i k \cdot x/N),
\]

(2)

where \( x \) and \( k \) are discretized in Eq. (2) and stand for pixels that range in each dimension from 0 to \( N - 1 \). Since only the magnitude of the Fourier transform can be experimentally measured, the data correspond to

\[
|F(k)| = \left| \sum_{x=0}^{N-1} f(x) \exp(2\pi i k \cdot x/N) \right|.
\]

(3)

Equation (3) is really a set of equations, and the phase problem is to solve these for \( f(x) \) at each pixel, where pixel means the element of the \( f(x) \) and \( F(k) \) arrays. Because of the loss of the phase, there are some ambiguities; that is, one cannot distinguish any of the following three quantities: \( f(x) \), \( f(x + x_0) \exp(i\theta) \), and \( f^*(-x + x_0) \times \exp(i\theta) \), where \( x_0 \), \( \theta \) are real constants and \( * \) donates complex conjugation. These three quantities are called the trivial characteristics of \( f(x) \). In the following, we study only the nontrivial characteristics of the phase problem of \( f(x) \). Let us discuss Eq. (3) under two conditions. First, we assume that \( f(x) \) is complex valued. For a 1D object the total number of equations of Eqs. (3) is \( N \), but the total number of unknown variables is \( 2N \) because each pixel has two unknown variables: the real part and the imaginary part. For 2D and 3D complex-valued objects the total number of equations is \( N^2 \) and \( N^3 \), respectively, and the total number of unknown variables is \( 2N^2 \) and \( 2N^3 \), respectively. Second, we consider \( f(x) \) real. According to Friedel’s law, the magnitude of its Fourier transform, \( |F(k)| \), has central symmetry. Therefore the equation number for a 1D real object drops to \( N/2 \), and the number of unknown variables is \( N \). For 2D and 3D real objects, the total number of equations is \( N^2/2 \) and \( N^3/2 \), respectively, and the total number of unknown variables is \( N^2 \) and \( N^3 \), respectively. On the basis of the above analysis, we suggest that, given the magnitude of a Fourier transform sampled at Bragg density, the phase problem is underdetermined by a factor of 2 for 1D, 2D, and 3D objects.

Apparent, without a priori information Eq. (3) cannot be solved, and the phase is not unique. To recover the phase information, we have to introduce some constraints on Eq. (3). The first strategy that we introduce is to decrease the number of unknown variables by using objects with some known scattering density (i.e., some known-valued pixels) inside them. A special example is to use an object with some nonscattering density (i.e., some zero-valued pixels) inside it. To determine how many known-valued pixels of \( f(x) \) are necessary for solving Eq. (3), we introduce the concept of ratio \( \sigma \), which is defined as

\[
\sigma = \frac{\text{total pixel number}}{\text{unknown-valued pixel number}}.
\]

(4)

where the unknown-valued pixels are to be solved for. Since, given the magnitude of a Fourier transform sampled at Bragg density, the phase problem is underdetermined by a factor of 2 for 1D, 2D, and 3D objects, the equations should be solvable, at least in principle, as long as the ratio \( \sigma > 2 \). One may argue that a necessary, but not sufficient, condition for finding a solution for Eq. (3) is that the number of unknown variables be equal to the number of equations. For example, since some of the equations could be linear combinations of others, it is possible that there are many solutions for Eq. (3) even though the number of equations is more than the number of unknown variables, but we know from Barakat and Newmas that this situation is pathologically rare. Furthermore, although Eqs. (3) are nonlinear, and each has two solutions because of the modulus sign, the positivity constraints (see Section 3) eliminate one of the two. This appears to be one of the important roles played by the positivity constraints, especially in the case of a loose support (see Section 4).

The second strategy to increase the number of known quantities of Eq. (3) is to use the oversampling method. The idea of oversampling is to sample the magnitude of a Fourier transform finely enough to get a finite support for the object in which the pixel value outside the finite support is zero. To solve the phase problem, we have to oversample the magnitude of the Fourier transform to make the ratio \( \sigma > 2 \). The requirement corresponds to the oversampling by \( > 2 \) for a 1D object, \( > 2^{1/2} \) in each dimension for a 2D square object, and \( > 2^{1/3} \) in each dimension for a 3D square object.

3. INTERNAL CONSTRAINTS

In Section 2 we discussed the finite support or zero-valued pixel constraints for phase retrieval, where we call these constraints the external constraints. In this section we will study the constraints on the object itself: the internal constraints.

As mentioned in Section 1, if an object is real and positive, one can use both the external and the internal (i.e., positivity) constraints to successfully recover the phase information from the magnitude of the Fourier transform. However, for complex-valued objects, the positivity constraints have been considered inapplicable, which has made phase retrieval difficult. However, they are not entirely inapplicable. In the case of X-ray diffraction, for example, the complex-valued object density can be expressed by using the complex atomic scattering factor, \( f_1 \)
1. Fourier transform an input object \( f_j(x) \) and get a Fourier pattern \( F_j(k) \) in reciprocal space.

2. Generate a new Fourier pattern \( F_{j+1}(k) \) by using \( |F(k)| \) as its magnitude and the phase of \( F_j(k) \) as its phase, \( F_{j+1}(k) = |F_j(k)| \times e^{i \arg F_j(k)} \).

3. Inverse Fourier transform \( F_{j+1}(k) \) to get a new object \( f_{j+1}(x) \).

4. Generate the \((j + 1)\)th object by

\[
f_{j+1}(x) = \begin{cases} f_j(x), & x, f_j(x) \in S \\ f_j(x) - \beta f_j^*(x), & x, f_j(x) \notin S \end{cases}
\]

where \( \beta \) is a constant between 0.5 and 1 and \( S \) represents those pixels that are inside a finite support and whose imaginary parts are positive. The lower part of Eq. (5) plays two roles: (i) if the pixels are outside the finite support, the pixel values decrease gradually to zero, and (ii) if the pixels are inside the finite support and their imaginary parts are negative, the imaginary parts of those pixels increase at every iteration until they are positive. This represents the modification to Fienup’s algorithm. For the \( j \)th iteration, a reconstruction error function \( E_j \) is introduced to monitor the progress of the algorithm,

\[
E_j = \left( \frac{\sum_{x \in S} |f_j(x)|^2}{\sum_{x \in \mathbb{R}^2} |f_j(x)|^2} \right)^{1/2}.
\]

In our computer phasing experiment, we generate the initial input by using the magnitude \( |F(k)| \) and a random phase. The constant \( \beta \) is set to 0.8. All of the objects that we try to reconstruct are complex valued, with real and imaginary parts. We mentioned in Section 3 that the real parts of complex-valued objects are usually positive and the imaginary parts are always positive; therefore we generate the original objects by setting all the imaginary parts and most of the real parts positive (approximately 90%) in the following computer phasing experiments.

The effect of the ratio \( \sigma \) on the quality of phase retrieval for 2D and 3D complex-valued objects was studied. First we investigated complex-valued objects in which
many zeros have been provided. Figure 1(a) illustrates
the magnitude of a complex object (all the figures in the
following show the magnitude of a complex-valued object)
of $512 \times 512$ pixels with a square hole of $\sigma = 4$ inside it
in which the pixel values are zero. The reason that we
set the square hole off center is to avoid any artificial ef-
fects of symmetry on the reconstruction. From the mag-
nitude of the object’s Fourier transform, which is not over-

![Image](image.png)

Fig. 3. Examples of image reconstruction from the magnitude of
the Fourier transforms of complex-valued objects by oversam-
pling; positivity constraints on the imaginary part were used to-
gether with loose support constraints. (a) Magnitude of an
original 2D complex-valued object. (b), (c), (d), and (e) Magni-
tude of the reconstructed objects with $\sigma = 4, 3, 2.6,$ and 2.5, re-
spectively. The “corner” symbols in (b)–(e) outline the region of
added zero pixels.
sampled, we retrieved the phase by using the constraints of zero-valued pixels inside the square hole (external constraints) and positivity constraints on the imaginary part of the object (internal constraints). Here we assume that we know the boundary between the zero-valued and the unknown-valued pixels. Figure 1(b) illustrates the successfully reconstructed result (although we display only the magnitude of the reconstructed object, we correctly reconstruct both the real and the imaginary parts of the object). Then we gradually reduced the ratio \( s \) to 3 and 2.6, respectively. The reconstructed objects still looked excellent, as Figs. 1(c) and 1(d) show. Curves 1B, 1C, and 1D of Fig. 2 show the reconstruction error versus iteration number for the reconstruction of Figs. 1(b), 1(c), and 1(d), respectively. From the three curves, one can see that the error function of the reconstruction stays at approximately 1% for a few thousand iterations and then suddenly drops to a much lower level. However, when the

Fig. 3. (continued). (f) Isodensity of the magnitude of an original 3D complex-valued object. (g), (h), (i), and (j) Isodensity of the magnitude of the reconstructed objects with \( s = 7.8, 4, 2.57, \) and 2.3, respectively.
ratio went down to 2.5, the algorithm hardly converged, as curve 1E of Fig. 2 illustrates, and the reconstructed object became noisy [Fig. 1(e)]. This result brings out an apparent contradiction between the theory of Section 2 and the computer phasing experiment. As we stated in Section 2, the phase of nonperiodic objects should be unique as long as the ratio $\sigma \geq 2$. But the phasing experiment showed that the phase can be perfectly retrieved only with the ratio $\sigma \geq 2.6$ for the case in Fig. 1. Our interpretation is that, when the ratio $\sigma$ is larger than a certain number (for the case in Fig. 1 the number is 2.6), the algorithm always converges to the global minimum and the correct phase information can always be recovered. However, when the ratio $\sigma$ is less than the number, the algorithm converges to a local minimum instead of the global minimum. In the following phasing experiments, we will see the same phenomenon.

We also used the oversampling method to generate zero-valued pixels around an object in which the nonperiodicity and the presence of an envelope permit us to provide those pixels. Figure 3(a) is an illustration of an original complex-valued object with $512 \times 512$ pixels. Figures 3(b), 3(c), and 3(d) are the successful reconstructions from the magnitude of the Fourier transforms with use of the positivity constraints on the imaginary part and loose supports with ratio $\sigma = 4$, 3, and 2.6, respectively, where a loose support is a support that is bigger than the true boundary of a specimen. In this example we choose as loose supports squares with edges four pixels larger than those of the square objects, and we define $\sigma$ as being equal to the ratio of the total pixel number to the pixel number inside the loose support. The reason that we tested with loose supports is that it may be difficult in actual practice to locate the true boundary of an object. However, when the ratio was reduced to 2.5, the reconstruction almost failed, as Fig. 3(e) shows. Curves 3B, 3C, 3D, and 3E of Fig. 4 illustrate the reconstruction error versus iteration number for the reconstruction of Figs. 3(b), 3(c), 3(d), and 3(e), respectively.

The reconstruction of 3D complex-valued objects from the magnitude of their Fourier transforms by using the oversampling method was also investigated. Figure 3(f) is an isodensity map of the magnitude of the original complex-valued object with $64 \times 64 \times 64$ pixels. Figures 3(g), 3(h), 3(i), and 3(j) were reconstructed from the magnitude of the Fourier transforms by using the positivity constraints on the imaginary part and loose supports with ratio $\sigma = 7.8$, 4, 2.57, and 2.3, respectively, where the loose supports are spheres with radii two pixels larger than those of the spherical objects. From these figures, one can see that the reconstruction is excellent, with $\sigma \geq 2.57$, but it failed with $\sigma = 2.3$. In Fig. 4, curves 3G, 3H, 3I, and 3J illustrate the reconstruction error versus iteration number for the reconstructions of Figs. 3(g), 3(h), 3(i), and 3(j), respectively. All the above computer phasing experiments showed that the phase is retrievable from the magnitude of the Fourier transform as long as the ratio is larger than a certain number. According to our computer phasing experience, this number depends somewhat on the object.

To examine the applicability of this approach to real data, we have to investigate the sensitivity of the algorithm to noise. In our computer phasing experiments, random noise

$$\text{noise} = \frac{\text{signal}}{\text{SNR}} \times \text{random} \quad (7)$$

was added to the magnitude of the Fourier transform, where SNR is the desired signal-to-noise ratio and the random function generates random numbers from −0.5 to 0.5. We used Fig. 3(a) as an original 2D complex-valued object. With added noise with SNR = 20, Fig. 5(a) was reconstructed by using positivity constraints on the imaginary part and a loose support with $\sigma = 2.6$. We then reduced SNR to 10 and kept the same loose support. Figure 5(b) illustrates the reconstructed object from which we are still able to see the features. The other factor that we considered is the central stop, which arises from the fact that the diffraction intensity (i.e., the absolute square of the magnitude of a Fourier transform) in the central pixels cannot be precisely measured by experiment. We modeled a central stop by removing the central $11 \times 11$ pixels from the magnitude of the Fourier transform, which corresponds to 0.05% of the total data. Figure 5(c) shows the reconstructed object with an $11 \times 11$ pixel central stop and SNR = 20. The effect of noise on the 3D object reconstruction was also studied. We used Fig. 3(f) as the original complex-valued 3D object. Figures 5(d) and 5(e) show two reconstructed isodensity maps of the magnitude of the original object with SNR = 40 and 20, respectively; we used positivity constraints on the imaginary part and loose supports with $\sigma = 2.57$. Meanwhile, we did not see any improvement of the reconstruction at larger values of $\sigma$. From Figs. 5(a), 5(b), 5(d), and 5(e) one can see that the reconstruction of the 3D complex-valued object with noise is not as good as that of 2D case. But this does not mean that 2D object reconstruction can tolerate much more noise than 3D object reconstruction. The reason for the difference is that in 2D reconstruction we illustrated the magnitude of complex-valued objects, but in 3D reconstruction we used an isodensity display, making a direct visual comparison somewhat difficult.

Fig. 4. Reconstruction error versus iteration number for the reconstruction of Fig. 3. Curves 3B, 3C, 3D, 3E, 3G, 3H, 3I, and 3J correspond to the reconstruction of Figs. 3(b), 3(c), 3(d), 3(e), 3(g), 3(h), 3(i), and 3(j), respectively.
5. SUMMARY

Our theoretical analysis suggests that, given the magnitude of a Fourier transform sampled at the Bragg density, the phase problem is underdetermined by a factor of 2 for 1D, 2D, and 3D objects. Therefore, at least in principle, the phase can be retrieved from the magnitude of a Fou-
rier transform as long as the ratio $\sigma > 2$, and the requirement of phase retrieval by oversampling the magnitude of a Fourier transform by $4 \times$ in 2D reconstruction and $8 \times$ in 3D reconstruction is unnecessary.

Having studied the physics of complex-valued objects, we propose that one can use the positivity constraints on the imaginary part as internal constraints for phase retrieval, thereby supplementing the effectiveness of the external constraints, especially when a loose support is employed. In the computer phasing experiments, we successfully reconstructed complex objects from the magnitude of their Fourier transforms with ratio $\sigma \geq 2.6$ for 2D and $\sigma \geq 2.57$ for 3D reconstruction by using positivity constraints on the imaginary part and loose supports. We also reconstructed complex-valued objects with reasonable noise and a central stop by using loose supports and positivity constraints on the imaginary part.

ACKNOWLEDGMENTS

The authors are grateful to Janos Kirz, professor in the Department of Physics, the State University of New York at Stony Brook, for many invaluable and stimulating discussions and to Chris Jacobsen, professor in the Department of Physics, the State University of New York at Stony Brook, for the use of his computer facility. This work was supported by the U.S. Department of Energy under grant DE-FG02-89ER60858.

Address correspondence to J. Miao at the address on the title page.

*Present address, 15 Jefferson Court, Bridgewater, New Jersey 08807.
*Present address, Advanced Microtechnology Program, Lawrence Livermore National Laboratory, Livermore, California 94551.

REFERENCES AND NOTES

17. A. Szoke, University of California, Lawrence Livermore National Laboratory, P.O. Box 808, L-41, Livermore, Calif. 94551 (personal communication, December 1995). Szoke said that he thought $2 \times$ oversampling should in theory suffice for any dimensionality $> 2$.
26. The exception is where the material amplifies the incident x-ray beam, as with the x-ray laser amplifier.