Multidimensional phase problems

R. P. Millane
Whistler Center for Carbohydrate Research and Computational Science and Engineering Program, Purdue University, West Lafayette, Indiana 47907-1160

Received May 8, 1995; revised manuscript received October 27, 1995; accepted October 31, 1995

Uniqueness properties of phase problems in three or more dimensions are investigated. It is shown that an N-dimensional image is overdetermined by its continuous Fourier amplitude. A sampling of the Fourier amplitude, with a density approximately $2^{2-N}$ times the Nyquist density, is derived that is sufficient to uniquely determine an N-dimensional image. Both continuous and discrete images are considered. Practical implications for phase retrieval in multidimensional imaging, particularly in crystallography where the amplitude data are undersampled, are described. Simulations of phase retrieval for two- and three-dimensional images illustrate the practical implications of the theoretical results.

© 1996 Optical Society of America

1. INTRODUCTION

There are many application areas in imaging in which one can accurately measure only the amplitude, but not the phase, of the Fourier transform of an image, rather than the image itself. Recovery of the image is then said to present a phase problem, since it is necessary to recover the phase in order to invert the Fourier transform to calculate the image.\textsuperscript{1–3} Phase problems usually occur either when a detected wave field suffers substantial phase distortion (resulting from propagation through a random medium, for example) or when the wavelength is too small to allow coherent detection. Examples occur, for example, in radio engineering, astronomy, microscopy, ultrasonic imaging, various forms of remote sensing, and nondestructive testing, radar imaging, and crystallography.\textsuperscript{1,3–7}

It is well known that one-dimensional phase problems do not have a unique solution; i.e., a one-dimensional image is not uniquely determined by the amplitude of its Fourier transform.\textsuperscript{8} However, in almost all practical situations two-dimensional phase problems do have a unique solution, and reasonably reliable reconstruction algorithms have been developed.\textsuperscript{3,4,9–16}

There are a number of scientific and technical applications of imaging that involve three- (and higher-) dimensional phase problems.\textsuperscript{3,7,17–19} These include crystallography, electron microscopy, nondestructive testing, ultrasonic imaging, geophysical imaging, radar imaging, spatiotemporal image processing, and analysis of higher-order spectra. Properties of these higher-dimensional phase problems are the subject of this paper. A preliminary report of some of these ideas has been presented.\textsuperscript{20}

Uniqueness properties of two-dimensional phase problems are summarized in Section 2. The main result is presented in Section 3, which is a proof that a three-dimensional continuous image is uniquely determined by a particular sampling of the Fourier amplitude that is at approximately half the Nyquist density. Extension of this result to N-dimensional images is outlined in Section 4. Analogous results for discrete images are outlined in Section 5. Implications of these results, particularly in crystallography, are described in Section 6. In Section 7 example simulations of image reconstruction for multidimensional discrete images are presented that illustrate the implications of the theory. Concluding remarks are made in the final section.

2. PROPERTIES OF TWO-DIMENSIONAL PHASE PROBLEMS

Consider a two-dimensional image \( f(x, y) \), which may be complex, that has compact support, i.e., is equal to zero outside a finite region of image space. The support of an image \( f(x, y) \), denoted by \( S[f(x, y)] \), is defined here as the smallest rectangular region in image space that contains the image. The dimensions of the support of \( f(x, y) \) are \( a \) and \( b \) in the \( x \) and \( y \) directions, respectively, and the support is taken here to be centered on the origin, so that

\[
S[f(x, y)] = (-a/2, a/2) \times (-b/2, b/2).
\] (1)

The Fourier transform of \( f(x, y) \) is denoted by \( F(u, v) \), so that

\[
F(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \exp[i2\pi(ux + vy)] \, dx \, dy.
\] (2)

The phase of \( F(u, v) \) is denoted by \( \tau[F(u, v)] \), i.e.,

\[
F(u, v) = |F(u, v)| \exp[i\tau[F(u, v)]],
\] (3)

where \( |F(u, v)| \) denotes the amplitude. The image is given by the inverse Fourier transform of \( F(u, v) \), i.e.,

\[
f(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(u, v) \exp[-i2\pi(ux + vy)] \, du \, dv.
\] (4)

The intensity \( |F(u, v)|^2 \) is the Fourier transform of the autocorrelation of the image, \( A(x, y) \), i.e.,

\[
|F(u, v)|^2 \leftrightarrow A(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x', y') f^*(x + x', y + y') \, dx' \, dy',
\] (5)

where \( \leftrightarrow \) denotes a Fourier-transform pair. The support of the autocorrelation is

\[
S[A(x, y)] = (-a, a) \times (-b, b)
\] (6)
and has twice the dimensions of the support of the image in each spatial dimension.

Although loss of the phase of $F(u, v)$ would appear, as a result of the form of Eq. (4), to preclude recovering the image $f(x, y)$ from $|F(u, v)|$, this is not the case because there are in fact few phase functions that are consistent with an image of compact support. The property of compact support is key to the uniqueness properties of two-dimensional phase problems. This is because the Fourier transform $F^*(\gamma)$ of such a function, continued into the two-dimensional complex space $\gamma = (u + i\xi, v + i\eta)$, is an entire function of $\gamma$, i.e., is analytic for all finite $\gamma$. It has been shown, by use of the zero sheets (the complex manifolds where $|F(\gamma)|^2$ vanishes), that the factors $F^*(\gamma)$ and $F^*(\gamma)$ can be separately determined from $|F(\gamma)|^2$, allowing $F(u, v)$, and thence $f(x, y)$, to be uniquely determined.8,12 As this involves analytically continuing $|F(u, v)|^2$ off the real $u-v$ plane into the complex space $\gamma$, uniqueness depends on $|F(u, v)|$ being known continuously in $u$ and $v$. Equivalently, as a result of the sampling theorem and expressions (5) and (6), $|F(u, v)|$ may be sampled with spacings no greater than $1/2a$ and $1/2b$ in the $u$ and $v$ directions, respectively. The data $|F(m/2a, n/2b)|$, for all integers $m$ and $n$, are sufficient to uniquely solve the two-dimensional phase problem. This set represents samples of the amplitude at the Nyquist density. It is worth pointing out that uniqueness occurs almost always, in the sense that exceptions occur with probability zero.11,12 These pathological cases are not considered in this paper.

Although the image is essentially uniquely determined by the Fourier amplitude, some characteristics of the image are irretrievably lost when the phase is lost.1,3 These are the absolute position of the image, a constant phase factor, and an ambiguity between the image and its inversion in the origin. What can be reconstructed, therefore, is $\hat{f}(x, y)$, where

$$\hat{f}(x, y) = f(sx - \alpha, sy - \beta \exp(i\phi)) \quad (7)$$

and $s = \pm 1$, $\alpha$, $\beta$, and $\phi$ are unknown real constants. However, these ambiguities are usually of little significance in practice. The corresponding ambiguities in the transform are

$$\hat{F}(u, v) = F(u, v) \exp[i(\phi + au + bv)] \quad \text{or}$$
$$F^*(u, v) \exp[i(\phi + au + bv)], \quad (8)$$

which can be written as

$$\hat{F}(u, v) = F(u, v) \exp\{i[kP[F(u, v) + \phi + au + bv]]\}, \quad (9)$$

where $k = s - 1 = 0$ or $-2$. The ambiguities in $F(u, v)$ result from the fact that $|\hat{F}(u, v)| = |F(u, v)|$ and that the inverse Fourier transform of $\hat{F}(u, v)$ has compact support with dimensions $a$ and $b$. The above description applies to continuous images; i.e., $x$ and $y$ are continuous variables. However, analogous results apply to discrete images $f[m, n]$, where $m$ and $n$ are integers, by using the analytic properties of the $z$ transform of the image.12,13,21,22 Discrete images are discussed in Section 5.

The uniqueness properties of two-dimensional phase problems summarized above are used in Section 3 to investigate the three-dimensional problem. Although the ambiguities expressed by Eqs. (7)–(9) are of little significance in the two-dimensional case, they are important in analyzing the three-dimensional case, as is shown below.

3. THREE-DIMENSIONAL PHASE PROBLEMS

Uniqueness properties of three-dimensional phase problems are derived in this section. Consider a three-dimensional image $f(x, y, z)$, which may be complex, that has compact support, with dimensions $a$, $b$, and $c$, i.e.,

$$S[f(x, y, z)] = (-a/2, a/2) \times (-b/2, b/2) \times (-c/2, c/2), \quad (10)$$

and Fourier transform $F(u, v, w)$. First consider the case where $|F(u, v, w)|$ is known for all $u$, $v$, and $w$. Because $|F(u, v, w)|$ is known continuously into the three-dimensional complex space $\gamma = (u + i\xi, v + i\eta, w + i\zeta)$. As in the two-dimensional case, because $|F(\gamma)|^2$ is analytic for finite $\gamma$ the factors $F(\gamma)$ and $F^*(\gamma)$ can be determined, allowing $F(u, v, w)$, and thence $f(x, y, z)$, to be calculated. Therefore, as in the two-dimensional case, the three-dimensional phase problem has a unique solution if the intensity is available continuously in Fourier space. Equivalently, the amplitude may be sampled at the Nyquist density; i.e., the amplitude samples $|F(m/2a, n/2b, p/2c)|$, for all integers $m$, $n$, and $p$, are also sufficient to uniquely determine the image.

In going from the one-dimensional case to the two-dimensional case, the phase problem goes from being nonunique to unique. One might suspect, therefore, that in going from two dimensions to three the problem may go from being unique to being “more than unique”: In other words, although $|F(u, v, w)|$ for all $u$, $v$, and $w$ is sufficient to uniquely define the image, it may not be necessary: fewer data, i.e., samples of $|F(u, v, w)|$ whose density is below the Nyquist density, may be sufficient. This is shown here to be the case by derivation of a sub-Nyquist sampling of the amplitude that uniquely defines the image. The approach is as follows. We consider $u-v$ planes in Fourier space that are separated by twice the Nyquist spacing in the $w$ direction. The amplitude on each of these planes is treated as representing a two-dimensional phase problem that can be solved uniquely for the phase on each plane. Phase problems on two additional planes are then solved that allow the solutions on the above set of planes to be used to reconstruct the complete complex amplitude $F(u, v, w)$, thus solving the three-dimensional phase problem. The required amplitude data are on planes that sample the continuous three-dimensional amplitude at almost half the Nyquist density. The proof is by construction; i.e., we prove uniqueness by showing how one could, in principle, construct the image uniquely from the given set of data.

Consider the Fourier transform $F(u, v, w)$ on the planes $w = p/c$ in Fourier space, for all integers $p$. The two-dimensional inverse Fourier transform of $F(u, v, p/c)$ is

$$f_p(x, y) = \int_{-\infty}^{\infty} f(x, y, z) \exp(i2\pi pz/c) dz. \quad (11)$$
This function has compact support with dimensions in the \(x\) and \(y\) directions that are no larger than those of \(f(x, y, z)\). Usually the support will be the same as that of \(f(x, y, z)\), i.e.,

\[
S[f_p(x, y)] = (-a/2, a/2) \times (-b/2, b/2).
\]

(12)

It is possible, however, for the support to be smaller than that of \(f(x, y, z)\). We show this by noting that \(f_p(x, y)\) is the \(p\)th coefficient of the Fourier series in \(z\) for \(f(x, y, z)\), i.e.,

\[
f(x, y, z) = (1/c) \sum_{p=-\infty}^{\infty} f_p(x, y) \exp(-i2\pi pz/c).
\]

(13)

If this coefficient vanishes over an appropriate region in \(x\) and \(y\), then the support of \(f_p(x, y)\) can be smaller than that of \(f(x, y, z)\). Although it is unlikely to occur in practice, this possibility is included in the analysis to keep the result general. Since \(f_p(x, y)\) has compact support with dimensions no greater than \(a\) and \(b\), the two-dimensional phase problem posed as recovering \(f_p(x, y)\) from \([F(u, v, p/c)]\) has a unique solution if \([F(u, v, p/c)]\) is available continuously in \(u\) or \(v\) or, equivalently, if the samples \(F(m/2a, n/2b, p/c)\) are available for all integers \(m\) and \(n\). However, as explained in Section 2, what is actually recovered is \(f'_p(x, y)\), where

\[
f'_p(x, y) = f_p(spx - \alpha_p, spy - \beta_p) \exp(i\phi_p)
\]

and \(s_p = \pm 1, \alpha_p, \beta_p,\) and \(\phi_p\) are unknown constants. Therefore \(\hat{F}_p(u, v)\), which is the two-dimensional Fourier transform of \(f'_p(x, y)\) given by

\[
\hat{F}_p(u, v) = \sum_{p=-\infty}^{\infty} f'_p(x, y) \exp(i[k_p\mathcal{P}[F(u, v, p/c)] + \phi_p + \alpha_p u + \beta_p v]),
\]

(15)

can, in principle, be determined from the data \([F(u, v, p/c)]\) for each \(p\). The goal is to build up \(F(u, v, w)\) from the \(F(u, v, p/c)\) so that the constants \(k_p = 0\) or \(-2\) and \(\alpha_p, \beta_p,\) and \(\phi_p\) need to be determined for each \(p\), for calculation of the \(F(u, v, p/c)\) from the \(\hat{F}_p(u, v)\), using Eq. (15). The strategy for doing this involves solving two additional two-dimensional phase problems on the planes \(u = 0\) and \(v = 0\) and matching the resulting solutions with \(\hat{F}(u, v, p/c)\) where these two planes intersect the planes \(w = p/c\), in order to determine the unknown constants.

The details are as follows.

Consider first solving the two-dimensional phase problem represented by the amplitude data \([F(0, v, w)]\) on the \(u = 0\) plane. As described above, the inverse Fourier transform of \(F(0, v, w)\) has a support that is no larger than that of \(f(x, y, z)\), so that either \([F(0, v, w)]\) or \([F(0, n/2b, p/2c)]\) for all integers \(n\) and \(p\), can be used to determine the quantity

\[
F(0, v, w) \exp(i[k\mathcal{P}[F(0, v, w)] + \phi + \beta v + \gamma w]).
\]

(16)

As \(f(x, y, z)\) can be determined only to within an inversion in the origin, a constant phase factor, and a shift in image space, arbitrary values can be chosen for \(k, \phi, \beta,\) and \(\gamma\), and then \(F(0, v, w)\) is known from expression (16).

Second, consider solving the two-dimensional phase problem represented by the amplitude data \([F(u, 0, w)]\) on the \(v = 0\) plane. From these data the quantity

\[
\hat{F}(u, 0, w) = F(u, 0, w) \exp(i[k\mathcal{P}[F(u, 0, w)] + \phi' + \alpha' u + \gamma' w])
\]

(17)

can be determined. The constant \(\alpha'\) can be set to an arbitrary value because it defines the absolute position of \(f(x, y, z)\) along the \(y\) direction. Substituting \(u = 0\) into Eq. (17) gives

\[
\hat{F}(0, 0, w) = F(0, 0, w) \exp(i[k\mathcal{P}[F(0, 0, w)] + \phi' + \gamma' w]),
\]

(18)

and \(F(0, 0, w)\) is known from \(F(0, v, w)\) [obtained from expression (16)] on \(v = 0\). Therefore the only unknowns in Eq. (18) are \(k', \phi',\) and \(\gamma',\) and these can be found by equating the \(w\) dependence of the left-hand side (lhs) and right-hand side (rhs) of this equation. Once these constants are determined, \(F(u, 0, w)\) can be calculated from \(F(u, 0, w)\) using Eq. (17).

The functions \(F(0, v, w)\) and \(F(u, 0, w)\) can now be used to determine the unknowns \(k_p, \phi_p, \alpha_p,\) and \(\beta_p\) in Eq. (15) for each \(p\) as follows: First, setting \(u = 0\) in Eq. (15) gives

\[
\hat{F}_p(0, v) = F(0, v, p/c) \exp(i[k\mathcal{P}[F(0, v, p/c)] + \phi_p + \beta_p v]),
\]

(19)

and, as \(F(0, v, p/c)\) is known, the unknowns \(k_p, \phi_p,\) and \(\beta_p\) can be determined by equating the \(v\) dependence of the lhs and rhs of Eq. (19). Second, setting \(v = 0\) in Eq. (15) gives

\[
\hat{F}_p(u, 0) = F(u, 0, p/c) \exp(i[k\mathcal{P}[F(u, 0, p/c)] + \phi_p + \alpha_p u]),
\]

(20)

and, as \(F(u, 0, p/c)\) is known, the remaining unknown \(\alpha_p\) can be determined by equating the \(u\) dependence of the lhs and rhs of Eq. (20). With the four constants determined, \(F(u, v, p/c)\) can be calculated from \(\hat{F}_p(u, v)\) by Eq. (15), for each value of \(p\). Because the complex amplitude \(F(u, v, p/c)\) is now known, by virtue of the sampling theorem, \(F(u, v, w)\) can be reconstructed for all \(w\), allowing \(f(x, y, z)\) to be calculated, and the phase problem is solved.

The above analysis shows that a three-dimensional image is (almost always) uniquely determined by the amplitude of its Fourier transform on a subspace \(T_3 \subset \mathbb{R}^3\) of the three-dimensional Fourier space \((u, v, w)\) given by

\[
T_3 = \{(u, v, p/c), (0, v, w), (u, 0, w)\}; \forall u, v, w \in \mathbb{R}, \forall p \in \mathbb{Z}\}
\]

(21)

rather than by the full space \((u, v, w); \forall u, v, w \in \mathbb{R}\), where \(\mathbb{R}\) and \(\mathbb{Z}\) denote the sets of real numbers and integers, respectively. The implications of this are more apparent if one considers the minimum set of samples, denoted by \(S_3\), that is sufficient, by the sampling theorem, to define the amplitude on \(T_3\), given by

\[
S_3 = \{(m/2a, n/2b, p/c), (0, n/2b, p/2c), (m/2a, 0, p/2c); \forall m, n, p \in \mathbb{Z}\}.
\]

(22)
This set contains one three-dimensional data set at half the Nyquist density and two two-dimensional data sets at the Nyquist density. The two-dimensional sets form a vanishingly small proportion of the total data, so the data that are sufficient to uniquely solve the phase problem are effectively at half the Nyquist density, i.e., effectively half of the Nyquist samples \((m/2a, n/2b, p/c); \forall m, n, p \in \mathbb{Z}\). The three-dimensional phase problem is then in a sense overdetermined by the fully sampled amplitude data \(F(u, v, w)\).

The analysis presented above accommodates the unusual cases in which the support of \(f_p(x, y)\) is smaller than that of \(f(x, y, z)\) for at least one value of \(p\). This is unlikely to occur in practice, and if these cases are excluded, then the number of sufficient amplitude data can be further reduced, as follows: If the support of \(f_p(x, y)\) has dimensions \(a \times b\) in the \(x\) and \(y\) directions, respectively, for all \(p\), then \(\alpha_p\) and \(\beta_p\) in Eq. (15) can be fixed by requiring that Eq. (12) be satisfied for each \(p\). The only unknowns in Eq. (15) are then \(k_p\) and \(\phi_p\), which can be determined by use of Eq. (19). Since Eq. (20) is then not needed, \(F(u, 0, p/c)\) is not needed, so the data \(F(u, 0, w)\) are not needed. Hence, in most practical cases, the data on the space denoted by \(T'_3\), where

\[
T'_3 = \{(u, v, p/c), (0, v, w); \forall u, v, w \in \mathbb{R}, \forall p \in \mathbb{Z}\},
\]

are sufficient to uniquely define the image. Furthermore, in many cases the image \(f(x, y, z)\) is positive, and this, or any other constraint, is likely to make the vanishingly small amount of data \(F(0, u, w)\) in Eq. (23) unnecessary, implying that the data on the space \(T''_3\), where

\[
T''_3 = \{(u, v, p/c); \forall u, v \in \mathbb{R}, \forall p \in \mathbb{Z}\},
\]

may be sufficient for us to uniquely define the image in many practical cases. The data on \(T''_3\) are strictly at half the Nyquist density.

Finally, it is important to note that the above description of the reconstruction of a three-dimensional image by solution of a set of two-dimensional phase problems is used here only as a means of proving uniqueness. Such an involved procedure would not be used in practice to actually reconstruct a three-dimensional image. Existing algorithms for two-dimensional phase retrieval\(^4,15,16\) are easily extended to higher dimensions.

4. **N-DIMENSIONAL PHASE PROBLEMS**

The approach used in Section 3 can be used to derive properties of \(N\)-dimensional phase problems for any finite \(N\). In Section 3 we derived properties of three-dimensional phase problems, using the properties of two-dimensional problems. In a similar way, the properties of \(N\)-dimensional phase problems can be derived from those of \((N-1)\)-dimensional problems. Beginning with \(N - 1 = 3\), the properties for any \(N\) can be obtained by induction. The four-dimensional case is outlined here.

Consider a four-dimensional image \(f(x, y, z, s)\) that has support with dimensions \(a, b, c, d\), and Fourier transform \(F(u, v, w, t)\). The amplitude \(|F(u, v, w, q/d)|\) can be treated as representing a three-dimensional phase problem for each integer \(q\) that we can solve by applying the results of Section 3, using amplitude data on the set

\[
\{(u, v, p/c, q/d), (0, v, w, q/d), (u, 0, w, q/d)\},
\]

to give \(F(u, v, w, q/d)\) within five unknown constants. Using the amplitude on the three (two-dimensional) planes \(u = v = 0, u = w = 0, v = w = 0\) and solving the corresponding two-dimensional phase problems allow these constants to be determined by matching the data \(F(u, v, w)\) and \(F(u, v, w)\) dependence of the solutions along the lines \((u = v = 0, t = n/d), (u = w = 0, t = n/d), (v = w = 0, t = n/d)\). Note that only two-dimensional, rather than three-dimensional, data are required at this point, because fixing the finite number of unknown constants the complex amplitude only along a line requires, and solution of a two-dimensional phase problem is sufficient for recovery of the necessary complex amplitude. Therefore amplitude data on the subspace \(T_4 \subset \mathbb{R}^4\), where

\[
T_4 = \{(u, v, p/c, q/d), (0, v, w, q/d), (u, 0, w, q/d), (0, 0, w, t); \forall u, v, w, t \in \mathbb{R}, \forall p, q \in \mathbb{Z}\}
\]

are sufficient to uniquely determine \(f(x, y, z, s)\). Each of the components of \(T_4\) in Eq. (26) can be sampled at the Nyquist density to give a sampled data set \(S_4\) analogous to Eq. (22) in three dimensions. This corresponds to one four-dimensional data set at one quarter of the Nyquist density, two three-dimensional data sets at half of the Nyquist density, and three two-dimensional data sets at the Nyquist density. As in the three-dimensional case, the lower-dimensional data form a vanishingly small number compared with the total, so that the amplitude data that are sufficient are effectively at one quarter of the Nyquist density.

As in the three-dimensional case, if the unusual images whose Fourier coefficients have smaller supports than does the image are excluded, the set \(T_4\) reduces to \(T'_4\), where

\[
T'_4 = \{(u, v, p/c, q/d), (0, v, w, q/d), (0, 0, w, t); \forall u, v, w, t \in \mathbb{R}, \forall p, q \in \mathbb{Z}\}
\]

In many practical cases in which other constraints such as positivity apply, this may effectively reduce to \(T''_4\), where

\[
T''_4 = \{(u, v, p/c, q/d); \forall u, v, \in \mathbb{R}, \forall p, q \in \mathbb{Z}\}
\]

so that the amplitude may be undersampled by a factor of 2 along each of two directions in Fourier space, resulting in undersampling strictly by a factor of 4 relative to the Nyquist density.

The general \(N\)-dimensional case can be treated similarly, showing that the amplitude data that are sufficient have a density of \(2^{2-N}\) times the Nyquist density. Hence a smaller proportion of the total amplitude data are required as the dimensionality increases. The exact amplitude data sufficient to uniquely define the image for higher-dimensional cases are easily derived but become rather unwieldy. In many practical cases it is probably sufficient if the amplitude is undersampled by a factor of 2, in all but two directions in Fourier space.
5. PHASE PROBLEMS FOR DISCRETE IMAGES

The analysis presented above applies to continuous images, i.e., $x$ and $y$ are continuous variables. Although most physical images are effectively continuous, it is usually necessary to sample an image, or its transform, to capture data and perform digital processing. Processing of sampled (discrete) data is performed by use of the discrete Fourier transform (DFT). It is therefore useful to examine the properties of multidimensional phase problems for discrete images. Such an analysis is presented in this section. Because the analysis follows that for continuous images rather closely, it is only outlined here.

A. Two-Dimensional Images

Consider an $M \times N$ two-dimensional image denoted by $f[m, n]$, where $m = 0, 1, \ldots, M - 1$ and $n = 0, 1, \ldots, N - 1$. It is convenient here to zero pad the image to fill a region of size $2M \times 2N$. The DFT of $f[m, n]$, $F[q, r]$, is then

$$F[q, r] = \sum_{m=0}^{2M-1} \sum_{n=0}^{2N-1} f[m, n] \exp[i2\pi (mq/2M + nr/2N)].$$

(29)

The inverse DFT is given by

$$f[m, n] = \frac{1}{4MN} \sum_{q'=0}^{2M-1} \sum_{r'=0}^{2N-1} F[q', r'] \exp[-i2\pi (mq/2M + nr/2N)].$$

(30)

The intensity $|F[q, r]|^2$ has an inverse DFT that is the discrete autocorrelation, $A[m, n]$, of $f[m, n]$. The support of the autocorrelation is the (discrete) region $[-M + 1, M - 1] \times [-N + 1, N - 1]$, which has dimensions $(2M - 1) \times (2N - 1)$. $|F[q, r]|^2$ calculated with Eq. (29) therefore represents $A[m, n]$ without aliasing. $F[q, r]$ could, in fact, be defined on a region one point narrower in each direction, but it is convenient here to use the larger region, and the difference is of little significance except for very small images.

Uniqueness of the phase-retrieval problem for two-dimensional discrete images has been established by a number of authors. In fact, uniqueness has been established more rigorously than for the continuous case by mapping of the DFT to a $z$ transform, which then takes the form of a polynomial in two variables. Multiple solutions to the phase problem can occur only if the $z$ transform corresponding to $|F[q, r]|^2$ can be factored (aside from the factors corresponding to $|F[q, r]|^2$ and $F'[q, r]$), i.e., is irreducible. Because the set of reducible polynomials has measure zero in the space of polynomials in two variables, uniqueness occurs almost always. The reader is referred to Refs. 9, 12, 13, and 23 for more detailed information on this topic.

As in the continuous case, loss of the phase leads to ambiguities in position, absolute phase, and inversion in the origin of the image. These ambiguities can be represented by four unknown constants, as in the continuous case.

B. Three-Dimensional Images

The approach to deriving properties of phase problems for three-dimensional discrete images is analogous to that for the continuous case. Two-dimensional phase problems are solved on planes in Fourier space, solutions on additional planes are used to fix the unknown constants, and the image can be reconstructed from a subset of the full set of amplitude data.

Consider an $M \times N \times P$ three-dimensional discrete image $f[m, n, p]$ that is zero padded to fill a $2M \times 2N \times 2P$ region, with DFT $F[q, r, s]$. Writing the intensity as a $z$ transform and extending the above uniqueness argument from two dimensions to three dimensions show that, because polynomials in three variables are generally irreducible, knowledge of the amplitude $|F[q, r, s]|$ at all $2M \times 2N \times 2P$ points almost always ensures unique specification of $f[m, n, p]$. However, as for the continuous case described in Section 3, fewer samples of the intensity are sufficient to uniquely define the image, as is shown here.

Consider the subset $|F[q, r, 2s]|$, where $q = 0, 1, \ldots, 2M - 1$, $r = 0, 1, \ldots, 2N - 1$, and $s = 0, 1, \ldots, P - 1$, of the intensity samples. The two-dimensional inverse DFT, with respect to $q$ and $r$, of $F[q, r, 2s]$ is a function supported on $[0, M - 1] \times [0, N - 1]$. The resulting two-dimensional phase problem can therefore be uniquely solved, except for a set of constants for each value of $s$. The constants can be determined (provided that $2M, 2N, 2P \geq 4$) by considering the two-dimensional phase problems represented by $|F[0, r, s]|$ and $|F[p, 0, s]|$ and equating the solutions, as in the continuous case. Therefore, the complex amplitudes $F[q, r, 2s]$ can be determined from the above subset of the intensity data. This is sufficient for calculating the image because

$$f[m, n, p] = \frac{1}{8MNP} \sum_{q=0}^{M-1} \sum_{r=0}^{N-1} \sum_{s=0}^{P-1} F[2q, 2r, 2s]$$

$$\times \exp[-i2\pi (mq/M + nr/N + ps/P)].$$

(31)

for $0 \leq m \leq M - 1, 0 \leq n \leq N - 1, 0 \leq p \leq P - 1$. Therefore the amplitude $|F[q, r, s]|$ measured on the set

$$[[q, r, 2s], [0, r, s'], [q, 0, s'], q = 0, 1, \ldots, 2M - 1, r = 0, 1, \ldots, 2N - 1, s = 0, 1, \ldots, P - 1, s' = 0, 1, \ldots, 2P - 1]$$

(32)

is sufficient for reconstructing the image uniquely. The number of samples in expression (32) is $\mathcal{N}_t = 4(MN + N + M)P$, compared to the minimum number $\mathcal{N}_z = (2M - 1)(2N - 1)(2P - 1)$ based on the minimum full set of amplitude samples. The ratio $\mathcal{N}_t/\mathcal{N}_z$ approaches one half for $M, N,$ and $P$ large, as in the continuous case. The result for three-dimensional discrete images is therefore analogous to that for three-dimensional continuous images. Analogous results also apply to higher-dimensional discrete images, showing that $2^{2N}$ of the full set of amplitude samples are sufficient to determine an $N$-dimensional discrete image.

6. IMPLICATIONS

The analysis presented above shows that a multidimensional image is overdetermined by the amplitude of its Fourier transform. A number of imaging problems in microscopy, ultrasonics, geophysics, radar imagery, and crystallography are inherently three dimensional. There
are two general situations in which the results presented here have useful implications.

The first concerns multidimensional phase problems when all of the Fourier amplitude data are available. Although phase-retrieval algorithms in two dimensions are reasonably reliable, difficulties can arise in convergence to the correct solution (stagnation), particularly if the amplitude data are noisy or if the image is complex. Since redundant data tend to ameliorate the effects of noise, it is likely that reconstruction algorithms will be more effective in the multidimensional case than in two dimensions. This means that stagnation may be less likely to occur, fewer iterations may be required, and the algorithm may be less sensitive to noise, missing data, or a complex image. Although uniqueness properties do not necessarily transpose to the performance of particular reconstruction algorithms, it is likely that a higher dimensionality will have a positive effect on performance.

The second, and more significant, implications of the results described here occur when not all of the amplitude data are available, i.e., when the amplitude is undersampled (relative to the Nyquist density). The important implication is that in such cases the amplitude data may, in fact, be sufficient to permit unique recovery of the image. The uniqueness results may also indicate what kinds of sampled data, or what kinds of supplemental information, are likely to render the solution unique. Undersampling of the amplitude data may be due to experimental limitations, noise, or detector limitations or malfunctions. One application that necessarily leads to considerable undersampling of the amplitude is x-ray (or electron or neutron) crystallography. Implications in crystallography are briefly outlined here. The discussion is not detailed, and the reader is referred to Refs. 3 and 24 for a more detailed discussion on the crystallographic phase problem and the reasons for, and implications of, undersampling.

In x-ray crystallography the measured data are the intensity of the Fourier transform of a three-dimensional periodic image \( f(x, y, z) \). What is to be reconstructed is one period, denoted by \( e(x, y, z) \), of this image. The image in this case is equal to the electron density of the molecule under investigation. From the electron density, the uniqueness results may also indicate what kinds of sampled data, or what kinds of supplemental information, are likely to render the solution unique. Undersampling of the amplitude data may be due to experimental limitations, noise, or detector limitations or malfunctions. One application that necessarily leads to considerable undersampling of the amplitude is x-ray (or electron or neutron) crystallography. Implications in crystallography are briefly outlined here. The discussion is not detailed, and the reader is referred to Refs. 3 and 24 for a more detailed discussion on the crystallographic phase problem and the reasons for, and implications of, undersampling.

In x-ray crystallography the measured data are the intensity of the Fourier transform of a three-dimensional periodic image \( f(x, y, z) \). What is to be reconstructed is one period, denoted by \( e(x, y, z) \), of this image. The image in this case is equal to the electron density of the molecule under investigation. From the electron density, the uniqueness results may also indicate what kinds of sampled data, or what kinds of supplemental information, are likely to render the solution unique. Undersampling of the amplitude data may be due to experimental limitations, noise, or detector limitations or malfunctions. One application that necessarily leads to considerable undersampling of the amplitude is x-ray (or electron or neutron) crystallography. Implications in crystallography are briefly outlined here. The discussion is not detailed, and the reader is referred to Refs. 3 and 24 for a more detailed discussion on the crystallographic phase problem and the reasons for, and implications of, undersampling.

One application of crystallography for which the properties derived here give a rather crisp result is a technique called fiber diffraction analysis.\(^2\)\(^3\)\(^2\) In this application the image \( f(x, y, z) \) is periodic along only one spatial dimension, say the z direction, with period c. The amplitude of the Fourier transform is therefore undersampled, by a factor of 2, only along the w direction in Fourier space. From the analysis presented here, such amplitude data would be sufficient to uniquely define the image. However, in a fiber diffraction experiment the diffracting specimen is disordered, so that what is measured is only the cylindrical average of \( |F(u, v, p/c)|^2 \) about the w axis, i.e., the quantity

\[
I_p(R) = \int_0^{2\pi} |F(R, \psi, p/c)|^2 d\psi,
\]

where \((R, \psi)\) are circular polar coordinates corresponding to the Cartesian coordinates \((u, v)\). Recovering \( f(x, y, z) \) from \( I_p(R) \) is clearly a very underdetermined problem. One solves this problem in practice by collecting additional data from supplemental experiments, by making use of information on other related molecular structures, or both, to recover \( F(u, v, p/c) \) from \( I_p(R) \).\(^3\)\(^2\)\(^6\)\(^2\)\(^7\) These methods are tedious, both experimentally and computationally, and are not always effective. What the results derived in this paper show is that it is necessary to recover only \( |F(u, v, p/c)| \), rather than \( F(u, v, p/c) \), from \( I_p(R) \) to uniquely define the image. Since it is not necessary to retrieve \( P[F(u, v, p/c)] \), fewer additional experimental data or other structural information would be required, lessening the experimental demands of this procedure.\(^2\)\(^5\)

It is worth noting that an approach similar to that described here has been used as the basis of a phase-retrieval algorithm for two-dimensional problems.\(^2\)\(^8\)\(^2\)\(^9\) This involved solving a set of one-dimensional phase problems on parallel lines in Fourier space and resolving the multiple ambiguities in each of these by matching the solutions along orthogonal lines. However, in the two-dimensional case, many orthogonal lines are needed for resolving the many ambiguities present in the one-dimensional problems.

7. EXAMPLES

As described in Section 6, the uniqueness results for multidimensional phase problems derived in this paper suggest that iterative reconstruction algorithms may perform better for higher-dimensional phase problems than for lower-dimensional problems. We examine this proposition in this section by reconstructing two- and three-dimensional discrete images from their fully sampled and undersampled Fourier amplitudes, using an iterative transform algorithm, and comparing the results.

The simulations were conducted as follows: discrete, positive real images were generated and zero padded to twice the number of samples in each spatial dimension, and the Fourier amplitudes were calculated with the fast Fourier transform. Random phases were assigned to the amplitudes, and the images were reconstructed by use of Fienup’s iterative reconstruction algorithms.\(^1\)\(^5\) Support and positivity constraints were applied in image space. To compare different simulations, a fixed protocol consisting of 20 cycles of error reduction (er), 100 cycles of
Table 1. Amplitude Data Used for Reconstruction of $M \times N$ and $M \times N \times P$ Images

<table>
<thead>
<tr>
<th>Case</th>
<th>Two Dimensions</th>
<th>Three Dimensions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$</td>
<td>F(q, r)</td>
</tr>
<tr>
<td>2</td>
<td>$</td>
<td>F(q, r, 2s)</td>
</tr>
<tr>
<td>3</td>
<td>$</td>
<td>F(q, 2r)</td>
</tr>
<tr>
<td>4</td>
<td>$</td>
<td>F(q, 2r)</td>
</tr>
</tbody>
</table>

Fig. 1. Error versus iteration number for reconstructions of a three-dimensional image, for cases (a) 1, (b) 2, (c) 3, (d) 4, as described in the text.

Fig. 2. Error versus iteration number for reconstructions of a two-dimensional image, for cases (a) 1, (b) 3, and (c) 4, as described in the text.
the data. For the undersampled cases, only those amplitudes at the sample points representing the undersampled amplitude data are replaced, and the values at the remaining sample points are left unchanged.

Several different sets of amplitude data were used in the simulations, and they are referred to as cases 1–4. Case 1 corresponds to the full set of amplitude data for both two- and three-dimensional images. Case 2 corresponds to the set $T_1$ described in Section 3 for a three-dimensional continuous image. There is no case 2 for a two-dimensional image. Case 3 corresponds to the set $T_3$ for a three-dimensional continuous image and to the set $\{(u, n/b), (0, v); \forall u, v \in \mathbb{R}, \forall n \in \mathbb{I}\}$ for a two-dimensional continuous image. Case 4 corresponds to undersampling the amplitude by a factor of 2 in one direction, i.e., to the set $T_{30}$ for a three-dimensional continuous image and to the set $\{u, n/b, u, v [R, n [I]\}$ for a two-dimensional continuous image. Cases 3 and 4 for a two-dimensional image were not discussed previously in this paper and are not expected to lead to a unique reconstruction. However, simulations were run for these cases to compare the effects of undersampling in two and three dimensions. The amplitude data used for these different cases, for discrete $M \times N$ and $M \times N \times P$ images, are summarized in Table 1.

Simulations were conducted for an $8 \times 8$ two-dimensional image and an $8 \times 8 \times 4$ three-dimensional image. Amplitudes were therefore calculated, and computations conducted, on $16 \times 16$ and $16 \times 16 \times 8$ grids. Results for the three-dimensional and two-dimensional images are presented in Figs. 1 and 2, respectively. The best reconstructed image (of those obtained from the five different starting phases) for each case is shown in Figs. 3 and 4 for the three- and two-dimensional images, respectively. The actual three-dimensional and two-dimensional images are shown in Figs. 3a and 4a, respectively. Results for fully sampled amplitude data (case 1) are shown in Figs. 1a and 2a. Essentially perfect reconstructions are obtained for the three-dimensional image (final error $e = e_{140} < 0.005$) for all starting phases. For the two-dimensional image the reconstructions are not as good, although four of the five are reasonable ($e < 0.06$). Furthermore, convergence is considerably more rapid for the three-dimensional image than for the two-dimensional image, as one can see by comparing Figs. 1a and 2a. This behavior is consistent with that expected from the theory, as a unique solution is expected in both cases; but the three-dimensional case is overdetermined relative to the two-dimensional case, and the algorithm is more effective for the former.

For the undersampled case 2 in three dimensions the error is shown in Fig. 1b. Inspection of the figure shows that three of the five reconstructions are quite good, with $e < 0.04$, and the remaining two are poor. In fact, the performance of the algorithm in this case is similar to that for the fully sampled two-dimensional case (Fig. 2a). This is consistent with theory developed in this paper in that the amplitude data should be just sufficient for uniquely reconstructing the images in both of these cases.

Results for case 3, for the three-dimensional and two-dimensional images, are shown in Figs. 1c and 2b, respectively. The reconstruction of the three-dimensional image is reasonable ($e < 0.08$) in two of the five cases, the

![Fig. 3. (a) Original and (b)–(e) reconstructed three-dimensional images for cases (b) 1, (c) 2, (d) 3, and (e) 4, as described in the text. Each row in the figure shows the four planes of the three-dimensional image.](image1)

![Fig. 4. (a) Original and (b)–(d) reconstructed two-dimensional images for cases (b) 1, (c) 3, and (d) 4, as described in the text.](image2)
progress is made toward reconstructing the image, the errors remaining essentially the same as those for the random starting phases. This is again consistent with the theory in that in the three-dimensional case the data are only slightly less than what is required for unique specification of the image, whereas in the two-dimensional case the data are considerably less than what is required.

Results for case 4 are shown in Figs. 1d and 2c. The algorithm makes no progress in reconstructing either the two-dimensional or the three-dimensional image. For this particular three-dimensional image, therefore, the available amplitude data appear to be insufficient to allow a unique reconstruction. Failure to reconstruct the two-dimensional image is as expected for this degree of undersampling.

Comparison of the best reconstructed images for each case (Figs. 3 and 4) with the original images (Figs. 3a and 4a) shows very good reconstructions for cases 1–3 in three dimensions and case 1 in two dimensions and very poor reconstructions for case 4 in three dimensions and for cases 3 and 4 in two dimensions. The visual quality of the reconstructions is therefore also consistent with the theory described.

Overall, then, performance of the iterative transform algorithm for two- and three-dimensional images tracks the uniqueness properties derived in this paper. This indicates the relevance of these results for practical image-recovery problems. In the examples presented here a limited number of iterations were used in the reconstruction algorithm, as the intent was to compare the results for different image dimensionalities and degrees of undersampling. In the cases in which good reconstructions were obtained (Figs. 1a–1c and 2a), inspection of the figures indicates that it may be possible to improve the poorer reconstructions by using more iterations or by adjusting the algorithm parameters. On the other hand, inspection of Figs. 1d, 2b, and 2c shows that adjustment of the algorithm is unlikely to improve its performance in these cases. Failure to reconstruct the three-dimensional image for case 4 (Fig. 1d) is likely to be due to the small size of the image used here. This is so because for a small discrete image the two planes of amplitude data removed in this case form a significant proportion of the total data. For a larger discrete image they form a smaller proportion of the total data, and their absence is expected to have a less significant effect.

8. CONCLUSIONS

The analysis presented in this paper shows that a multi-dimensional (three- or more-dimensional) image is overdetermined by the amplitude of its Fourier transform. This means that the continuous Fourier amplitude, or the Fourier amplitude sampled at the Nyquist rate, contains redundant information. A constructive proof shows that a particular sampling scheme for the amplitude, which has a density of approximately $2^2 - N$ times the Nyquist density, is sufficient to uniquely define an $N$-dimensional image. Extensions of the sampling theorem to the case of nonuniformly spaced samples at the Nyquist density therefore suggest that, in practice, other sampling patterns for the amplitude at the above density are also likely to be sufficient to allow unique reconstruction of the image. Since the continuous amplitude can be reconstructed from these samples, Nyquist sampling is not in fact necessary to fully sample a band-limited multi-dimensional amplitude distribution.

The implications of these results are twofold. First, they indicate that multidimensional images should be easier (with increasing dimensionality) to reconstruct from fully sampled (continuous) amplitude data. This may be important for some multidimensional imaging problems. Second, the results indicate the feasibility of reconstructing multidimensional images from undersampled amplitude data. This will be important in cases for which it is convenient, or necessary, to undersample the amplitude. It is particularly important in crystallography where the amplitude data are necessarily undersampled as a result of the physics of the experiment. In these applications the results provide a useful guide to how much, and what kind, of ancillary data are needed to uniquely define an image. Simulations using iterative phase-retrieval algorithms support both the theory presented here and its practical implications.

REFERENCES

17. B. C. McCallum and J. M. Rodenburg, “Simultaneous reconstruction of object and aperture functions from multiple