Reconstruction of an object from its symmetry-averaged diffraction pattern

Veit Elser[a]* and R. P. Millane[b]

[a]Department of Physics, Cornell University, Ithaca, NY 14853-2501, USA, and [b]Computational Imaging Group, Department of Electrical and Computer Engineering, University of Canterbury, Christchurch, New Zealand. Correspondence e-mail: ve10@cornell.edu

The problem of reconstructing an object from diffraction data that has been incoherently averaged over a discrete group of symmetries is considered. A necessary condition for such data to uniquely specify the object is derived in terms of the object support and the symmetry group. An algorithm is introduced for reconstructing objects from symmetry-averaged data and its use with simulations is demonstrated. The results demonstrate the feasibility of structure determination using a recent proposal for aligning molecules by means of their anisotropic dielectric interaction with an intense light field [Spence *et al.* (2005). *Acta Cryst.* A**61**, 237–245].

1. Introduction

If one is primarily interested in molecules, then crystals are merely a convenient alignment mechanism for improving the signal level in a diffraction experiment. The near-perfect alignment of molecules within a crystal, however, comes at a cost. In a crystal, alignment is achieved by introducing translational periodicity as well, and this restricts the sampling of the diffraction pattern to the crystal's reciprocal lattice. A more complete sampling of the diffraction pattern is advantageous for constraining the phases in the pattern (Millane, 1990; Sayre & Chapman, 1995). This has led to the suggestion of alternate alignment mechanisms (Spence & Doak, 2004), where not only would the molecules lack translational periodicity but, by arranging the molecular separations to be outside the range of coherence of the illuminating radiation, all intermolecular interference would be eliminated. If such an 'incoherent alignment' mechanism were realized, the scattered radiation would form the diffraction pattern of an individual molecule and be signal enhanced in proportion to the number of molecules in the illuminating beam.

One of the leading candidates for an incoherent alignment mechanism is the interaction of an elliptically polarized light field with the anisotropic molecular polarizability (Larsen *et al.*, 2000). As already noted (Spence *et al.*, 2005), this scheme suffers from a disadvantage that the energetics of the molecular orientation is degenerate with respect to a discrete symmetry group of order four generated by π rotations about the principal axes of the polarizability tensor. The scattered radiation from many molecules will therefore be an average of four, generally distinct, diffraction patterns, all related by π rotations. An incoherently averaged pattern will have no simple (linear) relationship to the scattering density: in particular, it does not correspond to the symmetry average of the scattering density.

Since an incoherently symmetry averaged diffraction pattern contains less information than the original pattern, the question of whether such data uniquely determine the scattering object is an important one. The sampling theorem places limits on the information contained in the diffraction pattern, and thus the size of the group that will yield to a reconstruction from incoherently averaged data will be limited as well. We obtain a practical uniqueness criterion in terms of the action of the group on the object support. When this criterion is satisfied, we show how the molecule's scattering density can be reconstructed from a symmetry-averaged diffraction pattern by a relatively straightforward modification of a projection-based phase-retrieval algorithm.

We note that the problem addressed here is related to other problems in crystallography where averaged diffraction data are measured. In particular, it is closely related to fiber diffraction analysis where, as a result of random rotations of the (oriented) scattering objects about an axis, one measures the cylindrical average of the diffracted intensity (Millane, 2001). The relationship to fiber diffraction is discussed in §5. We note also, as a matter of interest, that the problem we consider is in some sense opposite to the phenomenon that occurs with noncrystallographic symmetry in crystals, where the molecular symmetries effectively augment the undersampled (single-particle) diffraction pattern from the crystal (Millane, 1993).

In the next section, we determine a necessary criterion for uniqueness of reconstructions from symmetry-averaged diffraction data. In §3, we describe an algorithm for reconstructing the scatterer from such data and, in the following section, we illustrate its efficacy using simulations. The results are discussed in the final section in relation to phase-retrieval problems in crystallography in general and to fiber diffraction analysis.

2. Uniqueness

A first step in assessing the feasibility of reconstruction is to determine whether the object is uniquely specified by the available data in the absence of noise and without the benefit of additional *a priori* knowledge (positivity *etc.*) aside from knowledge of the support. For the usual phase problem, this has been extensively studied. For sampled diffraction data from crystals, uniqueness is related to the size of the overlap region of aliased autocorrelation supports (Millane, 1993). The latter depends on the size and shape of the molecular support (molecular envelope) and makes more precise previous estimates based on the ratio of the molecular volume to the unit-cell volume (Arnold & Rossmann, 1986). The presence of noncrystallographic symmetry also improves uniqueness (Millane, 1990, 1993). For continuous diffraction data from single particles, the phase problem has been shown to be overdetermined from knowledge of the complete continuous diffraction amplitude (Millane, 1996; Miao *et al.*, 1998). The data redundancy increases with dimensionality of the problem (Millane, 1996). Given these results, we might anticipate that some averaging of the diffraction data can be tolerated without loss of uniqueness. We consider here these issues for continuous diffraction data as a function of support region of the scatterer and when the diffraction pattern has been incoherently averaged with respect to a discrete group G.

We let $\rho(\mathbf{r})$ denote the scattering density of the object at the direct-space point $\mathbf{r}$ and $\tilde{\rho}(\mathbf{q})$ its Fourier transform, *i.e.* the diffraction pattern, where $\mathbf{q}$ is position in Fourier space. The primary source of data is the intensity of the diffraction pattern averaged over a discrete group G, *i.e.*

$$I(\mathbf{q}) = \sum_{g \in G} |\tilde{\rho}(g\mathbf{q})|^2. \tag{1}$$

Symmetry averaging reduces the information in the data since the intensity is constant on symmetry orbits: $I(\mathbf{q}) = I(g\mathbf{q})$ for all $g \in G$. The known or estimated object support, a set in direct space, is denoted by S.

In order to better see the relationship between the degrees of freedom in the reconstruction and the available data, we assume that the diffraction pattern is measured up to a maximum spatial frequency (resolution) that corresponds, after rescaling, to direct-space samples $\mathbf{r}$ on the integer grid $\mathbb{Z}^N$, where N is the dimensionality. The Fourier transform $\tilde{\rho}(\mathbf{q})$ is then spanned by the set of linearly independent basis functions $\exp\{i2\pi\mathbf{q} \cdot \mathbf{r}\}_{\mathbf{r} \in S}$, where the support S is a finite subset of $\mathbb{Z}^N$ and $\mathbf{q}$ is the continuous transform variable on the N-torus. In this context, the sampling theorem is the statement that the coefficients of the Fourier transform basis are uniquely determined by the values of the Fourier transform at a set of $|S|$ samples, where $|S|$ is both the size of the basis and the area (volume) in pixels of the support. The locations of the Fourier samples are largely arbitrary, although numerical stability is improved by homogeneous distributions on the torus.

The diffraction intensity, $|\tilde{\rho}(\mathbf{q})|^2$, is spanned by the basis $\exp\{i2\pi\mathbf{q} \cdot \mathbf{r}\}_{\mathbf{r} \in S_A}$, where $S_A = S - S$ is the support of the autocorrelation of $\rho(\mathbf{r})$. The notation $S - S$ denotes the set of all difference vectors between points in S (the Minkowski sum of S and $-S$, where $-S$ is the inversion of S through the origin). The coefficients of the basis functions for $\{\mathbf{r}, -\mathbf{r}\} \in S_A$ are not independent, being related by complex conjugation. From this we infer that the diffraction intensity is completely specified by $|S - S|/2$ coefficients that can be determined, by the sampling theorem, from the same number of samples. In the case of complex-valued objects, where the autocorrelation coefficients are complex, the real intensities at $\mathbf{q}$ and $-\mathbf{q}$ represent two independent real-valued samples. On the other hand, if the object and autocorrelation are real-valued, then the data provided by a Friedel pair collapses to a single real number (since the intensities are equal). We can summarize both cases, complex or real objects, by the statement that an object comprising $|S|$ independent (complex/real) coefficients is constrained through its diffraction intensity to have a particular set of $|S - S|/2$ autocorrelation coefficients of the same type.

We first consider the question of uniqueness in the absence of symmetry averaging. If object support and diffraction intensity are the only constraints on the reconstruction, then a necessary condition for uniqueness is that the number of constraints exceeds the number of free parameters. From the remarks above, this criterion corresponds to the inequality $|S - S|/2 > |S|$, since constraints and free parameters are either both complex or both real. When this inequality is violated, then multiple reconstructions are consistent with the available data.

A useful measure of the sufficiency of the constraints in a reconstruction is the *constraint ratio*, Ω, where

$$\Omega = \frac{\text{number of independent autocorrelation coefficients}}{\text{number of independent object coefficients}}$$
$$= \frac{|S - S|}{2|S|}. \tag{2}$$

Values of Ω less than 1 or only slightly greater than 1 are cause for concern that the reconstruction may not succeed without additional constraints (*e.g.* positivity). On the other hand, if Ω exceeds 1, then the degree of nonuniqueness in a reconstruction will be very limited. In the usual case, this includes only translation or inversion of the object within the support (for loose or centrosymmetric supports) and a global phase rotation of the values. For pathological instances (*e.g.* homometric objects), there can be additional forms of nonuniqueness when $\Omega > 1$. As a practical matter, we can expect that the performance of reconstruction algorithms is improved, all other things being the same, when Ω is large.

It is well known that reconstructions in one dimension ($N = 1$), with one connected support, are not unique. This is an instance of the marginal case $\Omega = 1$, where almost any additional constraint (*e.g.* positivity) will restore uniqueness. A direct consequence of the Brunn–Minkowski theorem (Matousek, 2002), applied to subsets of $\mathbb{R}^N$ corresponding to the discrete sets S and $-S$, is the inequality

$$\Omega = \frac{|S - S|}{2|S|} \geq 2^{N-1}, \tag{3}$$

where equality holds when S is convex and centrosymmetric ($S = -S$). It has been observed that reconstructions are easier when $S \neq -S$ (Crimmins, 1987) or when S is not convex, as in a support with separated components (Fienup, 1987). This supports our contention that Ω is relevant for the performance of algorithms. In two dimensions, for example, a rectangular or elliptical support has $\Omega = 2$, while any triangle has $\Omega = 3$. Likewise, if support S_1 has Ω_1 as its constraint ratio, then two sufficiently separated copies S_2 of S_1 has $\Omega_2 = (3/2)\Omega_1$. For n sufficiently separated copies of S, without any symmetry relationships between their separations, $\Omega_n = (n - 1 + 1/n)\Omega_1$. Finally, Ω can be arbitrarily large, as in the case of a support with the shape of a thin circular (spherical) shell, for which the autocorrelation support is the whole interior of a circle (sphere), or for the n separated supports described above when $n \rightarrow \infty$. We note that, except in rather unusual cases, the molecular envelopes of globular proteins and other macromolecules tend to be approximately convex and centrosymmetric, and therefore in these cases $\Omega \approx 2^{N-1}$, or $\Omega \approx 4$ in three dimensions.

The effect of symmetry averaging is to reduce the number of independent autocorrelation coefficients, and thereby reduce the constraint ratio Ω. We examine two situations: the case of supports that can discriminate between some elements of the group orbit ($S \neq gS$ for some $g \in G$), and the more likely situation, where the support is invariant with respect to the action of G, either because the object itself is highly symmetric or the estimates of the support shape are imprecise.

The first situation is best explained using an example. In Fig. 1, we consider a rectangular support with dimensions a and b ($b \geq a$), which already possesses ambiguity with respect to inversion through the origin, denoted by i. As an example of G, we take the group generated by the mirror m

Figure 1
A rectangular support (left) and symmetry averaging by a mirror at 45° with respect to the edges of the support rectangle. The autocorrelation support, after symmetry averaging, is shown on the right (union of two doubled rectangles). Independent autocorrelation coefficients are represented by the shaded region; all others are related by the group of order four generated by two orthogonal mirrors (dashed lines). If the dimensions of the support rectangle are a and b ($b > a$), then the area of the shaded region is $a(2b - a)$. Dividing this by the area of the support, ab, we obtain $\Omega = 2 - a/b$.

oriented at 45° with respect to the rectangle edges. Since the support is not invariant with respect to m, the support constraint selects the two-elements orbit generated by i. Also shown in Fig. 1 is the support of the autocorrelation and its image under G. The pixels in the shaded region represent the independent autocorrelation coefficients that can be measured by sampling the symmetry-averaged diffraction pattern. By taking the ratio of the area of this region (independent data), $a(2b - a)$, and the area of the support (free parameters), we obtain $\Omega = 2 - a/b$. When $b > a$, it should be possible to reconstruct the object without additional constraints since $\Omega > 1$. The situation is better for a larger aspect ratio, although Ω is bounded by $\Omega < 2$, *i.e.* the value for any rectangle without averaging. For $b = a$, or square support, we have $\Omega = 1$ and uniqueness is marginal, *i.e.* not unique in practice in the presence of noise and without additional constraints. In the latter case, the averaging has reduced Ω from 2 to 1.

Now consider the situation where the support is invariant with respect to the group G. By counting the number of independent coefficients in the support of the autocorrelation, reduced by the size of the orbits under G, and forming the ratio with the number of free parameters in the support of the object, we obtain

$$\Omega = \frac{|S - S|}{|\langle i, G \rangle||S|}, \qquad (4)$$

where $|\langle i, G \rangle|$ is the order of the group generated by inversion in the origin and G [equation (3) is recovered when $\langle i, G \rangle = \langle i \rangle$]. This sets a limit on the size of the group that can be tolerated in a reconstruction from symmetry-averaged data. Note that, even in the case of uniqueness, the solution will be unique only up to a $|\langle i, G \rangle|$-fold ambiguity of symmetry-related objects. The least favorable situation arises for convex centrosymmetric supports since these saturate the Brunn–Minkowski bound (3), in which case, using (3) and (4), uniqueness requires that

$$|\langle i, G \rangle| < 2^N. \qquad (5)$$

When G includes i, the uniqueness criterion reduces to

$$|G| < 2^N, \qquad (6)$$

i.e. less than eightfold averaging in the three-dimensional case. If G does not include i, then $|\langle i, G \rangle| = 2|G|$ and uniqueness requires

$$|G| < 2^{N-1}, \qquad (7)$$

i.e. less than fourfold averaging in the three-dimensional case. The presence of additional *a priori* information will relax these conditions.

3. Reconstruction algorithm

In cases for which the degree of averaging admits a unique solution, we now address how to find that solution. For this purpose, we use the difference map algorithm (Elser, 2003*a,b*), a general-purpose iterative method for finding a point in a Euclidean space that lies in the intersection of two constraint
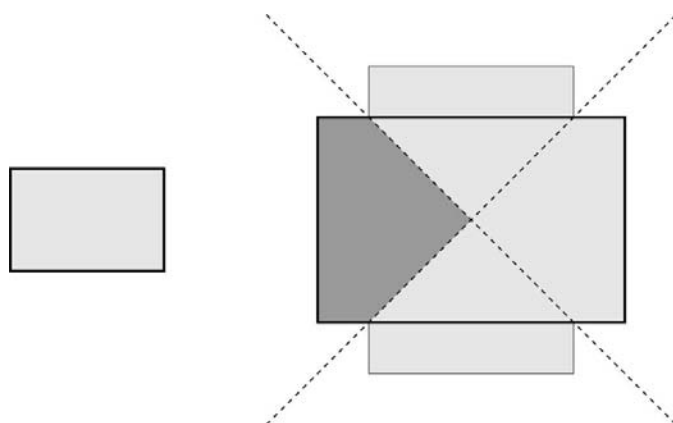
sets. The key components of the algorithm are the two projections that, given an arbitrary input, return as outputs the nearest points on the corresponding constraint sets.

The first projection implements the support constraint:

$$P_S : \rho(\mathbf{r}) \mapsto \rho'(\mathbf{r}) = \begin{cases} 0 & \text{if } r \notin S, \\ \rho(\mathbf{r}) & \text{otherwise,} \end{cases} \quad (8)$$

where S is the known or estimated support. Positivity (of real-valued objects), if it applies, is imposed by the projection $P_+(\rho(\mathbf{r})) = \max(0, \rho(\mathbf{r}))$. Since the support and positivity projections commute, they may be combined into a single projection P_{S+}. Although we consider only support and positivity constraints here, other direct-space constraints are easily incorporated within the same formalism.

The second projection, P_I, implements the constraint provided by the diffraction intensity and is simple when expressed in terms of the Fourier transform of the scattering density, $\tilde{\rho}(\mathbf{q})$. The projection computation therefore takes the form

$$P_I = F^{-1}\tilde{P}_I F, \quad (9)$$

where F is the discrete Fourier transform from the direct-space grid to the Fourier space grid. The projection $\tilde{P}_I$ acts on $\tilde{\rho}(\mathbf{q})$ and its distance-minimizing property in Fourier space extends to the direct-space density because F is a distance-preserving unitary transformation. Our main task is to determine the form of $\tilde{P}_I$ in the case of symmetry-averaged diffraction data.

Since we have a discrete group G that averages the diffraction intensity, we consider the action of $\tilde{P}_I$ on a symmetry orbit $\{\tilde{\rho}(g\mathbf{q})\}_{g \in G}$:

$$\tilde{P}_I : \tilde{\rho}(g\mathbf{q}) \mapsto \tilde{\rho}'(g\mathbf{q}). \quad (10)$$

For almost all $\mathbf{q}$, the size of the symmetry orbit will be the order of the group, $|G|$. The constraint is given by

$$\sum_{g \in G} |\tilde{\rho}'(g\mathbf{q})|^2 = I(\mathbf{q}), \quad (11)$$

where $I(\mathbf{q})$ is the symmetry-averaged intensity. The output of $\tilde{P}_I$ satisfies (11) while also minimizing the distance

$$\sum_{g \in G} |\tilde{\rho}(g\mathbf{q}) - \tilde{\rho}'(g\mathbf{q})|^2. \quad (12)$$

The set of input numbers $\{\tilde{\rho}(g\mathbf{q})\}_{g \in G}$ is a vector in a $2|G|$-dimensional real vector space and the geometrical content of (11) and (12) corresponds to finding the nearest vector $\{\tilde{\rho}'(g\mathbf{q})\}_{g \in G}$ that lies on a $(2|G| - 1)$-dimensional sphere of radius $[I(\mathbf{q})]^{1/2}$. The required projection is thus a simple rescaling of the input vector:

$$\tilde{P}_I(\tilde{\rho}(g\mathbf{q})) = \tilde{\rho}'(g\mathbf{q}) = \left[\frac{I(\mathbf{q})}{\sum_{g \in G} |\tilde{\rho}(g\mathbf{q})|^2} \right]^{1/2} \tilde{\rho}(g\mathbf{q}). \quad (13)$$

We now use the two projections P_S and P_I in the difference map D with $\beta = 1$ (see Elser, 2003a), for which it reduces to a generalization of Fienup's hybrid input–output map (Fienup, 1982),

$$D : \rho \mapsto \rho' = \rho + \Delta(\rho), \quad (14)$$

where

$$\Delta(\rho) = P_S(2P_I(\rho) - \rho) - P_I(\rho). \quad (15)$$

Application of (14) to an object ρ produces a new object ρ' and represents one iteration of the reconstruction algorithm. Repeated iterations drive the algorithm towards a fixed point of D where $\Delta(\rho) = 0$. Once arrived at a fixed point, the reconstruction is successful since the intensity estimate,

$$\rho_I = P_I(\rho), \quad (16)$$

is consistent with the support estimate

$$\rho_S = P_S(2P_I(\rho) - \rho). \quad (17)$$

The detailed form of D (for arbitrary β) is designed to maximize the attraction of the fixed points (Elser, 2003a). As a result of noise, of course, the algorithm will not converge exactly to a fixed point but, if $\|\Delta(\rho)\|$ is small, then $\rho_I \approx \rho_S$ and either one can be taken as the best estimate of the object.
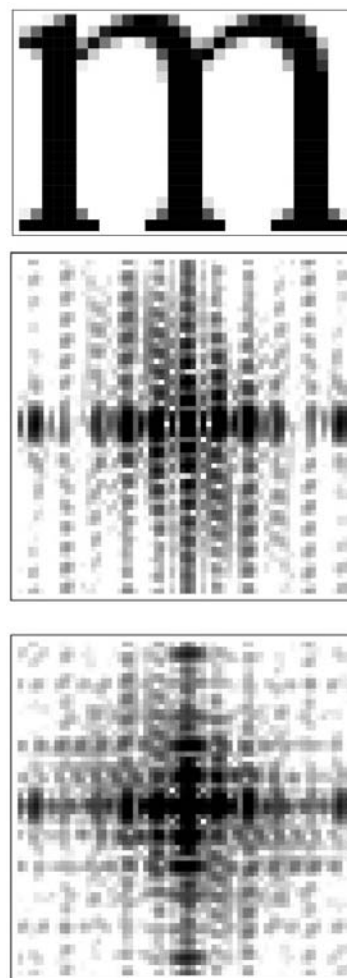

Figure 2
Top: object of study in a two-dimensional reconstruction experiment. Middle: diffraction pattern. Bottom: diffraction pattern averaged with respect to a diagonal mirror, as in Fig. 1. Diffraction patterns are displayed on a log scale.

If the problem has a unique solution, then this reconstruction will correspond to the true object. Of course, if the problem is not unique, then the algorithm may find a solution that satisfies the constraints but does not correspond to the true object.

It is worth pointing out that ρ itself is *not* an estimate of the solution, a mistake that has been made by a number of investigators.
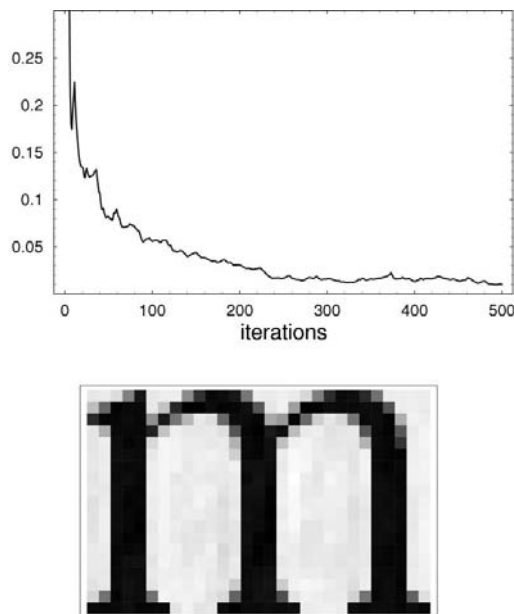

Figure 3
Top: Difference map error metric $\|\Delta\|$ *versus* iteration. Bottom: Object reconstructed from symmetry-averaged data.

4. Numerical experiments

We present here numerical simulations for two cases that illustrate the uniqueness results and the reconstruction algorithm described above. The first is in two dimensions and for the case where the object support is not invariant under the action of the group G. The second is in three dimensions and for the case where the object support is invariant with respect to G. Since the point of these exercises is to demonstrate the effects of symmetry averaging, we did not allow for noise, missing central data and other experimentally relevant factors. We believe that these factors will only compromise the results with ideal conditions when the constraint ratio is close to marginal ($\Omega \approx 1$).

The object of the first experiment, a non-negative letter 'm' on a 19×29 pixel support, is shown in Fig. 2. The object was embedded on a 64×64 grid and its diffraction pattern calculated, also on a 64×64 grid, and is shown together with its symmetry average with respect to a diagonal mirror in Fig. 2. This example corresponds to the situation described in §2 for a rectangular support that is not invariant under the action of the group generated by the mirror. The calculation of the constraint ratio is shown in the caption to Fig. 2 and gives $\Omega \approx 1.33$. Since $\Omega > 1$, there is sufficient information for a unique reconstruction. It is essential that the diffraction intensity is sampled at sufficient density that the number of symmetry-independent samples equals, or preferably exceeds, the number of independent autocorrelation coefficients. In this case, $(64 \times 64/4) = 1024 > (1.33 \times 19 \times 29) = 732$, so that this condition is satisfied with a ratio of ~ 1.4. Reconstruction was attempted using the symmetry-modified intensity constraint projection (13), and the standard support projection without positivity. The difference-map error metric $\|\Delta\|$ *versus* iteration is shown in Fig. 3. Convergence was obtained after about 500 iterations and the reconstruction is
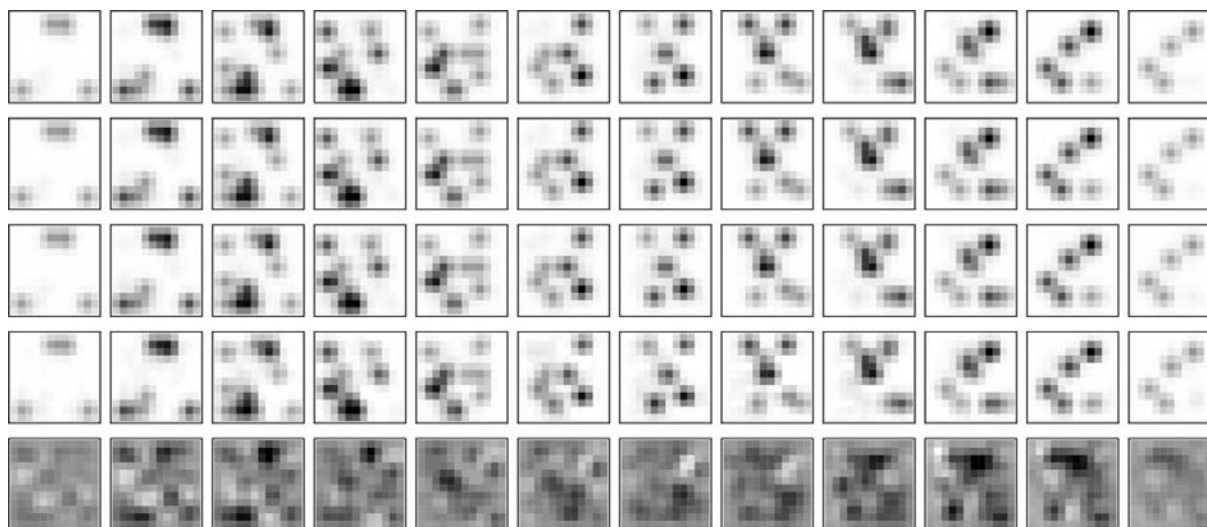

Figure 4
A three-dimensional object (top row) and its reconstructions (lower rows). Each horizontal row of 12, 12×12 pixel arrays, represents one three-dimensional density sampled on a 12^3 grid. The four reconstructions correspond to the conditions given in Fig. 5. The second row is for $|G| = 1$, the third for $|G| = 2$, the fourth for $|G| = 4$ with positivity, and the last row for $|G| = 4$ without positivity.

also shown in Fig. 3. In this example, the support cannot discriminate between the object and its inversion through the origin, and an essentially perfect reconstruction is obtained. Repeated application of the algorithm with many different random starts all converged to the correct object.

The second experiment mimics the symmetry ambiguity of the molecular alignment mechanism proposed by Spence *et al.* (2005). A non-negative three-dimensional density was generated by placing 25 unit 'atoms' at random positions within a 10^3 cube of zeros. After applying a low-pass filter, the resulting density had a support on a 12^3 cube and is shown in the top row of Fig. 4. The object was embedded on a 32^3 grid and its diffraction pattern calculated, also on a 32^3 grid, as well as its averages with respect to groups of order $|G| = 2$ and $|G| = 4$ generated by π rotations about axes along the sampling grid. When combined with the Friedel symmetry i of real-valued objects, the intensity data thus had a symmetry group $\langle i, G \rangle$ whose order ranged from 2 (no averaging) to 8 (three orthogonal π rotations). The latter, together with the choice of support, represents a marginal case, $|S - S|/|S| = 8$, and therefore by (4) we obtain $\Omega = 1$. Without additional constraints, the reconstruction may not be unique and, in any case, will be impossible in the presence of noise. As a check that the intensity was sampled at sufficient density even in this case, we note that the number of independent data, $(32^3/8) = 4096$, exceeds the number of independent autocorrelation coefficients, $(1 \times 12^3) = 1728$, by a ratio of ~ 2.4.

Reconstructions were obtained with the difference-map algorithm for the different data sets. For the cases $|G| < 4$ ($\Omega > 1$), the reconstructions were consistently successful (unique). Since the case $|G| = 4$ is marginal ($\Omega = 1$), an additional constraint in the form of positivity was included in the reconstruction algorithm and this consistently produced the correct solution. Reconstructions with $|G| = 4$ and no

positivity constraint converged to incorrect solutions, as expected. Fig. 5 shows the difference-map error metric $\|\Delta\|$ *versus* iteration for the four different cases. The corresponding reconstructions are compared with the correct object in Fig. 4 and it is seen that they are correct in the first three cases and incorrect in the fourth. Convergence is rapid for the most overdetermined case and slows as the order of the symmetry-averaging increases. Small values of $\|\Delta\|$ are ultimately obtained in all cases. For the marginal case ($|G| = 4$) without positivity, convergence is relatively rapid since the algorithm can quickly find one of a multitude of (non-unique) solutions. This example emphasizes the essential role of direct-space constraints in addition to support (*e.g.* positivity) for the scheme proposed by Spence *et al.* (2005).

5. Discussion

Although the phase-retrieval problem in macromolecular crystallography is underdetermined, requiring additional information (isomorphous derivatives, anomalous scattering, noncrystallographic symmetry) or constraints (positivity, atomicity) for its solution, the problem for continuous diffraction from single particles is known to be over-determined (Millane, 1996; Miao *et al.*, 1998). The work described here extends the latter results to quantitatively describe redundancy for general support (molecular envelope) shapes. Redundancy, as quantified by the constraint ratio Ω, increases as the support shape moves away from being convex and centrosymmetric. If the measured diffraction pattern is averaged over a discrete group of symmetries, which can occur when the specimen contains molecules that adopt a finite number of different orientations, then the loss of information due to the averaging is offset against the redundancy associated with the support. This relationship is made precise
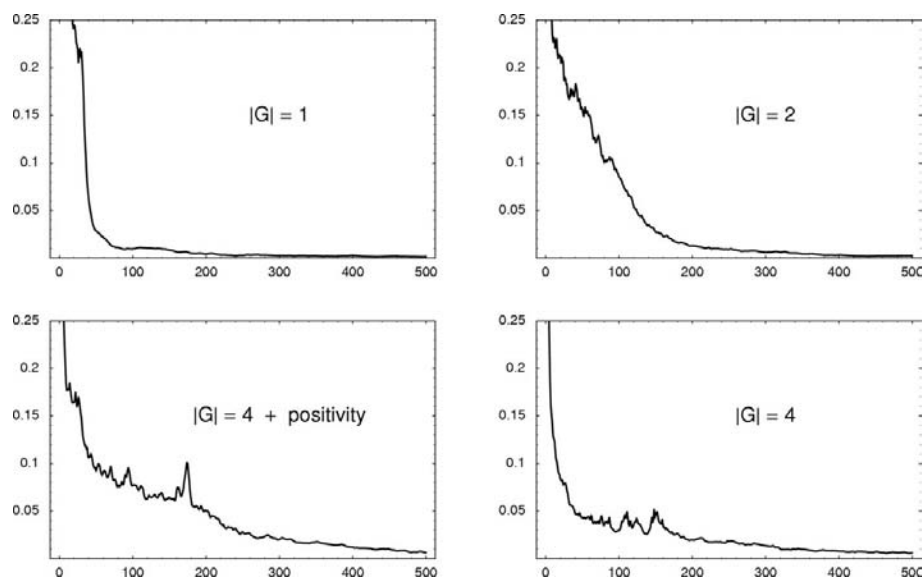

Figure 5
Difference-map error metric $\|\Delta\|$ *versus* iteration for the three-dimensional reconstructions with increasing symmetry averaging. The corresponding reconstructions are shown in Fig. 4.

in terms of the support and the symmetry group, giving conditions for uniqueness. For the supports expected in protein crystallography, less than fourfold or eightfold symmetry will be tolerated, depending on whether or not the group includes inversion through the origin. Additional *a priori* information will relax these requirements.

An implementation of the difference-map reconstruction algorithm that incorporates a projection operator onto the set of symmetry-averaged diffraction intensities is developed. Application of this algorithm to small numerical problems without noise leads to unique solutions. This supports the uniqueness arguments and the algorithm is a practical tool for determining structures from such data. In particular, this indicates that data collected from the laser alignment scheme proposed by Spence *et al.* (2005) should be sufficient to determine protein structures with minimal additional *a priori* information.

The constraint ratio Ω supersedes previous measures of redundancy for the phase problem with continuous data. Earlier work by Millane (1996), specialized for cuboid supports, gave the same dimensionality dependence but was off by a factor of 2. Miao *et al.* (1998) introduced the 'oversampling ratio' σ as, effectively, the number of intensity samples divided by the size of the object as measured in pixels. This combines in one number the redundancy intrinsic to the object support, Ω, with the degree to which the number of intensity samples exceed the sampling theorem minimum. In the case of low-noise data, where sampling on grids finer than the sampling theorem minimum does not add information, the value of σ is misleading in that a large value does not imply the reconstruction will be any easier.

The problem addressed in this paper is related to other problems in crystallography where averaged diffraction data are measured. In particular, it is closely related to fiber diffraction analysis where, as a result of random rotations of the (oriented) scattering objects about an axis, one measures the cylindrical average of the diffracted intensity (Millane, 2001). This corresponds to the action of a continuous rotation group in the formulation presented here. Despite the infinite order of the continuous group, fiber diffraction problems are quite tractable for the following reasons. First, in the case of polycrystalline fibers where *crystallites* are randomly rotated, the continuous group acts on sampled diffraction data so that at finite resolution it effectively reduces to a discrete group of finite order. Second, for noncrystalline fibers, where individual *molecules* are randomly rotated, the molecules generally have screw (helix) symmetry and at finite resolution the diffraction data can be expanded in a finite set of basis functions and the action of the continuous group again effectively reduces to a discrete group of finite order. The order of the groups encountered in fiber diffraction is modest, but is generally too large for a unique solution from the diffraction data alone, and augmentation of the data using multidimensional isomorphous replacement and/or molecular replacement is required (Millane, 2001). However, in particular, the implications of having access to continuous diffraction from noncrystalline fibers for reducing the amount of ancillary data required has been noted (see Section 6 of Millane, 1996). We also note that the projection operator (13) is identical to the operations used in density modification, molecular replacement and difference Fourier synthesis in fiber diffraction analysis, where one has estimates of the individual complex amplitudes and the measured averaged intensity (Namba & Stubbs, 1987; Baskaran & Millane, 1999; Millane, 2001).

References

Arnold, E. & Rossmann, M. G. (1986). *Proc. Natl Acad. Sci. USA*, **83**, 5489–5493.
Baskaran, S. & Millane, R. P. (1999). *IEEE Trans. Image Process.* **8**, 1420–1434.
Crimmins, T. R. (1987). *J. Opt. Soc. Am. A*, **4**, 124–134.
Elser, V. (2003a). *J. Opt. Soc. Am. A*, **20**, 40–55.
Elser, V. (2003b). *Acta Cryst.* A**59**, 201–209.
Fienup, J. R. (1982). *Appl. Opt.* **21**, 2758–2769.
Fienup, J. R. (1987). *J. Opt. Soc. Am. A*, **4**, 118–124.
Larsen, J. J., Hald, K., Bjerre, N., Stapelfeldt, H. & Seideman, T. (2000). *Phys. Rev. Lett.* **85**, 2470–2473.
Matousek, J. (2002). *Lectures on Discrete Geometry. Graduate Texts in Mathematics*, Vol. 212, p. 297. Berlin: Springer.
Miao, J., Sayre, D. & Chapman, H. N. (1998). *J. Opt. Soc. Am. A*, **15**, 1662–1669.
Millane, R. P. (1990). *J. Opt. Soc. Am. A*, **7**, 394–411.
Millane, R. P. (1993). *J. Opt. Soc. Am. A*, **10**, 1037–1045.
Millane, R. P. (1996). *J. Opt. Soc. Am. A*, **13**, 725–734.
Millane, R. P. (2001). *International Tables for Crystallography*, Vol. B, 2nd ed., edited by U. Shmueli, pp. 446–481. Dordrecht: Kluwer Academic Publishers.
Namba, K. & Stubbs, G. (1987). *Acta Cryst.* A**43**, 64–69.
Sayre, D. & Chapman, H. N. (1995). *Acta Cryst.* A**51**, 237–252.
Spence, J. C. H. & Doak, R. B. (2004). *Phys. Rev. Lett.* **92**, 198102–198104.
Spence, J. C. H., Schmidt, K., Wu, J. S., Hembree, G., Weierstall, U., Doak, B. & Fromme, P. (2005). *Acta Cryst.* A**61**, 237–245.