Acta Cryst. (1999). A**55**, 601–612

A feasible set approach to the crystallographic phase problem

L. D. Marks,* W. Sinkler and E. Landree

Department of Materials Science and Engineering, Northwestern University, Evanston, IL 60208, USA. E-mail: l-marks@nwu.edu

(*Received 2 June 1998; accepted 4 November 1998*)

I ask you to look both ways. For the road to a knowledge of the stars leads through the atom; and important knowledge of the atom has been reached through the stars.

Sir Arthur Stanley Eddington, 1882–1944†

Abstract

The connection between the crystallographic phase problem and the feasible set approach is explored. It is argued that solving the crystallographic phase problem is formally equivalent to a feasible set problem using a statistical operator interpretable *via* a log-likelihood functional, projection onto the non-convex set of experimental structure factors coupled with a phase-extension constraint and mapping onto atomic positions. In no way does this disagree with or dispute any of the existing statistical relationships available in the literature; instead it expands understanding of how the algorithms work. Making this connection opens the door to the application of a number of well developed mathematical tools in functional analysis. Furthermore, a number of known results in image recovery can be exploited both to optimize existing algorithms and to develop new and improved algorithms.

1. Introduction

X-ray diffraction and transmission electron diffraction both measure the intensity in the diffraction plane, from which the modulus can be extracted but not the phase – the crystallographic phase problem. Determining approximate values for the phase using what are called direct methods is now well established (Woolfson, 1987; Woolfson & Fan, 1995; Gilmore, 1996; Bricogne, 1984; Sheldrick, 1990; Giacovazzo, 1980; Dorset, 1996), sufficiently so that the structure of many small molecules can be solved relatively straightforwardly. In general, direct methods employ probabilistic approaches, derived from the fact that the scattering comes from atoms, to establish connections among the phases. Following this, peaks in the resulting charge density maps can be associated with atoms, and further recycling of the phases and fragments of the structure used in structure completion steps until all the atoms are located to within 0.1 Å when full structural refinements are carried out.

The crystallographic phase problem is by no means the only occurrence of a phase problem; it arises in numerous other areas as diverse as electron microscopy, wave-front sensing, interferometric imaging in astronomy (Gerchberg, 1974; Gerchberg & Saxton, 1972; Fienup, 1978; Dainty & Fienup, 1987) and measurement of the current *versus* magnetic field across a Josephson Junction (Dynes & Fulton, 1971). Developments in these areas have been along somewhat different lines, more related to the formal mathematics of a general image recovery problem. These methods have a history (and literature) as extensive as that for the crystallographic phase problem, and can be traced back at least as far as the 1930's (Kaczmarz, 1937). To quote from an article by Combettes (1996):

> Four basic elements are required to solve an image recovery problem:
> 1. A data formation model:
> 2. *A priori* information:
> 3. A recovery criterion:
> 4. A solution method.

These four criteria apply equally well to the crystallographic phase problem: we know quantitatively the formula connecting the scattering and the atomic positions; we (may) know the number of atoms and we know that the charge density is positive; we have probabilistic criteria for the recovery and iterative solution methods.

This raises the question of whether the crystallographic phase problem and the image recovery problem are really that different. In terms of how the problems have been approached algorithmically, crystallographers are used to dealing with very large numbers of reflections and exploit multisolution methods, while the image recovery problem has focused around mathematical techniques using what are called 'convex sets' (Sezan, 1992; Combettes, 1996; Combettes & Trussell, 1990; Gubin *et al.*, 1967), and iterative methods whose formal convergence behavior can be

† Sir Arthur Stanley Eddington (1927). In *Stars and Atoms*. New Haven: Yale University Press; London: H. Milford/Oxford University Press

analyzed in detail. The mathematics of convex projections is a well established analytical approach to the general problem of image recovery. It currently has a number of technologically important applications, for example the phase problem in interferometric images for astronomy, as well as all fields involving computer-aided tomography (Herman, 1980).

The intention of this note is to show that there is a very large amount of common ground between image recovery methods using what is called the feasible set approach and the crystallographic phase problem. In no way does this disagree with or dispute any of the existing statistical relationships available in the literature; instead it expands understanding of how the algorithms work. Making this connection opens the door to the application of a number of well developed mathematical tools in functional analysis. Furthermore, a number of known results in image recovery can be exploited both to optimize existing algorithms and to develop new and improved algorithms.

The structure of this paper is as follows. The next section (§2) outlines some of the basic elements of the feasible set approach, primarily definition of a number of terms such as sets, convexity, operators and projections, as well as some of the known results. The following section (§3) makes the connection between existing direct methods, structure completion techniques and the feasible set approach. A number of simple examples, not by any means intended to be exhaustive in scope, are then given (§4). Finally (§5), some possible extensions, such as using parallel statistical relationships converted to functionals or operators, and one suggestion about how to incorporate measurement errors are discussed, as well as a few mathematical issues which need further research.

2. The feasible set approach

We will provide here a brief introduction to the feasible set approach, as a prelude to the later sections of the paper. While in certain cases the formal mathematical shorthand of the literature will be used, this will be restricted as much as possible; this note is not being written for mathematicians. A useful overview of the concepts involved is provided in an article by Sezan (1992) and more formal analyses are available in the articles in a book edited by Stark (1987) and the recent review article by Combettes (1996; hereinafter referred to as PLC).

To start, we will be dealing almost exclusively with the case for which we know the moduli of the structure factors in reciprocal space and want to obtain phase estimates so that we can recover an approximation to the real-space charge density and, from this, approximate atomic coordinates for subsequent refinement. Structure factors, whether they are the true ones, unitary or normalized, will be written in terms of a generalized structure factor $\mathbf{X}$ unless otherwise specified. The Fourier transform of $\mathbf{X}$ will be written as $\mathbf{x}$, standing for a generalized charge density. Both $\mathbf{X}$ and $\mathbf{x}$ are vectors in a Hilbert space, $\mathbf{X}$ with complex values, $\mathbf{x}$ being only real. For later reference, we will consider that the complex vector $\mathbf{X}$ can be written as

$$\mathbf{X} = |\mathbf{X}| \exp(i\boldsymbol{\varphi}), \tag{1}$$

where $|\mathbf{X}|$ is a real positive vector, the values of which are known from the experimental data, and $\boldsymbol{\varphi}$ is a vector of the unknown phases. We will now define a number of terms that are required for later stages of the analysis.

We specify those values of $\mathbf{X}$ (or $\mathbf{x}$) which have certain properties as belonging to some set, S. In standard shorthand, this is written as

$$\mathbf{X} \in S. \tag{2}$$

For any vector in the Hilbert space, we can define a magnitude or norm of any vector $\mathbf{X}$ as

$$||\mathbf{X}|| = \left(\sum_i X_i^* X_i \right)^{1/2} \tag{3}$$

and a 'metric' corresponding to a distance between two vectors $\mathbf{X}$ and $\mathbf{Y}$:

$$\mathrm{d}(\mathbf{X}, \mathbf{Y}) = ||\mathbf{X} - \mathbf{Y}||. \tag{4}$$

[Some of the metrics used later in the paper are not those normally associated with a Hilbert space, suggesting that it would be better to use the more general term 'Banach space' (*e.g.* see Ramkrishna & Amundson, 1985). However, this is perhaps more a technical mathematical detail, not a substantive issue for the main results of this paper.] For any given set, a particularly important question is whether it is 'convex'. For any two members $\mathbf{X}$ and $\mathbf{Y}$ of a given set, the set is convex if a third point $\mathbf{Z}$ is also a member, where

$$\mathbf{Z} = \lambda\mathbf{X} + (1 - \lambda)\mathbf{Y}, \quad 0 < \lambda < 1, \tag{5}$$

i.e. all points lying on the line connecting $\mathbf{X}$ and $\mathbf{Y}$ belong to the set. Also of relevance later is the idea of a 'functional', loosely defined as a function of either $\mathbf{X}$ or $\mathbf{x}$ (as appropriate) which gives a real-valued scalar or vector, although we will only use scalars herein. The sections of some functional $g(\mathbf{X})$, *i.e.* the region

$$g(\mathbf{X}) < \beta \tag{6}$$

with β a scalar, are also a set. Using the same notation as above, a functional is convex (as is the associated set for arbitrary β) if

$$g[\lambda\mathbf{X} + (1 - \lambda)\mathbf{Y}] \leq \lambda g(\mathbf{X}) + (1 - \lambda) g(\mathbf{Y}), \quad 0 < \lambda < 1. \tag{7}$$

Finally, consider some general operator T which acts on $\mathbf{X}$ to give some new point, *i.e.*

$$T(\mathbf{X}) = \mathbf{Z}, \tag{8}$$

where $\mathbf{Z}$ is a modified structure factor. [Similarly, $T(\mathbf{x})$ will give a modified charge density.] We can also consider as a set the eigenvectors of T, known as the fixed points (Fix T) of the operator. An operator is contractive if

$$||T(\mathbf{X}) - T(\mathbf{Y})|| \leq k||\mathbf{X} - \mathbf{Y}||, \quad 0 < k < 1, \tag{9}$$

and nonexpansive if

$$||T(\mathbf{X}) - T(\mathbf{Y})|| \leq ||\mathbf{X} - \mathbf{Y}||. \tag{10}$$

If the operator is contractive, the set of fixed points is convex and there is a unique fixed point (Youla, 1987; PLC, pp. 168–170). If the operator is non-expansive then the fixed point is not unique, but it nevertheless can be proven under certain conditions that the set of fixed points is convex (Youla, 1987). A more general class of operator of importance to direct methods is not rigorously nonexpansive (though it may behave as a contraction for regions of the Hilbert space; see below). In such cases, the set of fixed points may be nonconvex and discontinuous. An important consequence of nonexpansivity is found by taking

$$Y = T(\mathbf{X}) \tag{11}$$

so

$$||T(\mathbf{X}) - T(T(\mathbf{X}))|| \leq ||\mathbf{X} - T(\mathbf{X})||. \tag{12}$$

Thus, for a nonexpansive operator T, $T(\mathbf{X})$ is always closer or the same distance from a fixed point of T than is $\mathbf{X}$.

We now need to define what is meant by a 'projection'. Suppose we have some set S and some point $\mathbf{X}$ which is not a member of the set. Let $\mathbf{Y}$ be the point on S such that $||\mathbf{X} - \mathbf{Y}||$ is minimized. The projection of $\mathbf{X}$ onto the set S, written as $P(\mathbf{X})$, is equivalent to

$$P(\mathbf{X}) = \mathbf{Y}. \tag{13}$$

Projections typically arise associated with constraints, and it is convenient to refer to a constraint that leads to a convex set as a 'convex constraint'. Finally, the concept of a mapping (not to be confused with the crystallographic terminology of map for an estimated $\mathbf{x}$) converts from one Hilbert (or Banach) space to another. For instance, a Fourier transformation is a mapping from $\mathbf{x}$ to $\mathbf{X}$.

We now have enough of the preliminaries to state the feasible set problem in a concrete fashion. Suppose that we want to consider finding the 'best' value of some scalar functional $g(\mathbf{X})$ (*e.g.* a figure of merit or FOM) which is consistent with a number of constraints. Each of these constraints can be represented as a set S_i, $i = 1, \ldots, n$, either convex or nonconvex. The standard optimization approach is

$$\text{minimize } g(\mathbf{X}) \quad \text{subject to} \quad \mathbf{X} \in S_1 \cap \ldots S_n, \tag{14}$$

where $\cap$ is the standard notation for intersection of sets, *i.e.* find the minimum value that obeys all the constraints. Rather than solving this particular problem, we associate a set S_o with the values of $g(\mathbf{X})$ which are below some certain value [*e.g.* equation (6) above], so the problem is to find the set of points $\mathbf{X}$ at the intersection of all the sets S_i, $i = 0, \ldots, n$, *i.e.*

$$\mathbf{X} \in S_0 \cap S_1 \cap \ldots S_n. \tag{15}$$

All such values are reasonable solutions to the problem. We can in fact define as our set of feasible solutions the intersection of all the sets S_i, *i.e.*

$$S_f = S_0 \cap S_1 \cap \ldots S_n. \tag{16}$$

Note that the feasible set depends both upon the constraints and how in detail we define the set S_0. While we want this set to be convex and continuous, there is no reason *a priori* why it should not be nonconvex and discontinuous. (In practice, we may want to start with a relatively loose constraint on S_0 and tighten it as the calculation proceeds.)

The key results are (Youla, 1987; PLC) as follows.

(i) If all the sets are convex, there exists one and only one possible intersection for all the sets; see *e.g.* Fig. 1.

(ii) If all the sets are convex, and for each set S_i we can associate a projection operator $P^i(\mathbf{X})$, then the sequence

$$X_{m+1} = X_m + \lambda(P^n(P^{n-1}(\ldots P^0(\mathbf{X}_m))) - \mathbf{X}_m), \; 0 < \lambda < 2, \tag{17}$$

where m is the iteration number, converges to a point in the unique solution set S_f [provided S_f is not empty; see point (iii)]. In other words, applying consecutive projections for convex sets finds the unique feasible set S_f. Here λ is a relaxation parameter, which for best results is typically in the range $1 < \lambda < 2$ (Levi & Stark, 1987; PLC, p. 168). Fig. 2 illustrates how the algorithm works assuming two convex sets and different values of the relaxation parameter.

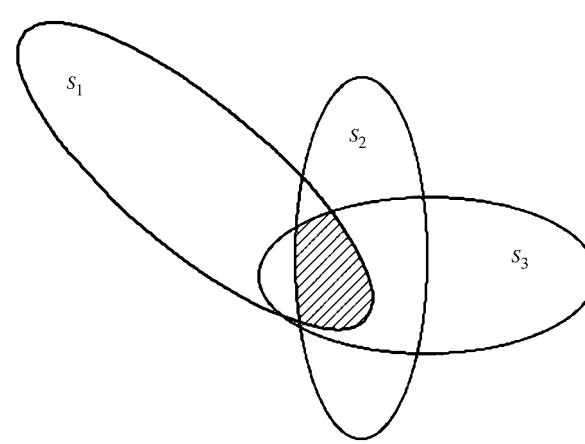

Fig. 1. Intersection of three convex sets. The shaded region corresponds to the feasible set of solutions.

(iii) If the constraint on the FOM is too tight, for instance there are experimental measurement errors or some of the constraints are only approximately (or probabilistically) correct, there may be no true solution as the union of the sets. However, we can still consider optimal solutions based upon minimizing the distance of a point from all the sets (PLC, pp. 202–209; Chrétien & Bondon, 1996).

(iv) If one or more of the sets is not convex, there may be more than one local solution and no unique solution *via* the successive projection approach of equation (17). However, the approach is locally convergent (PLC, p. 186), so will yield local minima of the problem considered as a feasible set [point (iii) above].

(v) If instead of using a projection we employ a contractive or nonexpansive operator $T(\mathbf{X})$, the fixed points of which lie within the set $g(\mathbf{X}) < \beta$, then since this is a downhill step towards a fixed point we can replace one or more of the projections in (17) by operators, as illustrated in Fig. 3. Hence, the iteration of (17) with operators always yields a better or equal point, converging if all the sets are convex in a global fashion. According to the literature, the method will converge locally in general if one or more of the sets are not convex, although we will see below that this depends upon when the FOM is calculated (see also Fig. 4). For later reference, we note that, near to any fixed point, $||T(\mathbf{X}) - \mathbf{X}||$ can always be expanded as a quadratic, so the region around any fixed point (local minimum) is necessarily locally convex.

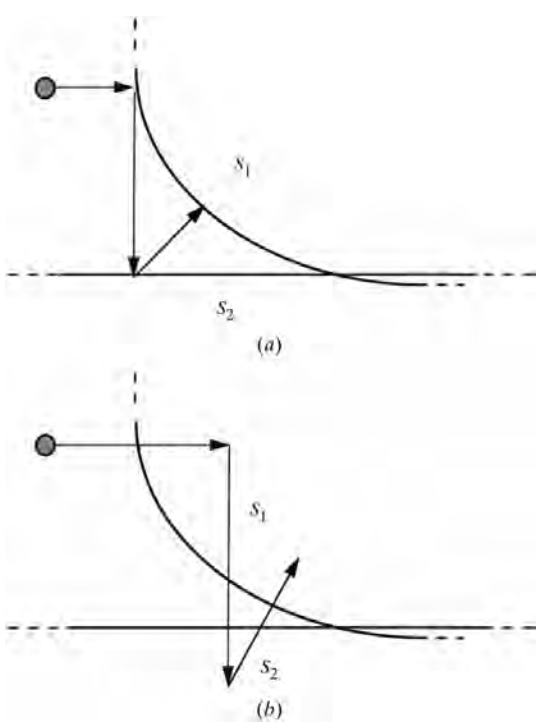

Fig. 2. Examples of (*a*) an unrelaxed projection operator, $\lambda = 1.0$, and (*b*) an over-relaxed projection operator, $\lambda > 1.0$.

3. Direct methods as a feasible set problem

Having now defined the basics of the feasible set method, we will show how it can be used in a very general sense for direct methods. We have available a set of experimental $|\mathbf{X}|$ values. This provides a constraint and a corresponding set within which lie all possible values for the phase, *i.e.* the set of complex numbers $\mathbf{X}$ which have the appropriate moduli and arbitrary phase. This is a nonconvex set, as may be verified from equation (5). Associated with any value $\mathbf{X}$ is a mapping (Fourier transform) to an associated charge density $\mathbf{x}$. For well measured data to finite resolution, the true charge density is real and positive and the set of all real positive $\mathbf{x}$ is a convex set. The final set of relevance is all the possible atomic coordinates for a finite number of atoms that can be obtained by some mapping (*e.g.* peak location) from $\mathbf{x}$.

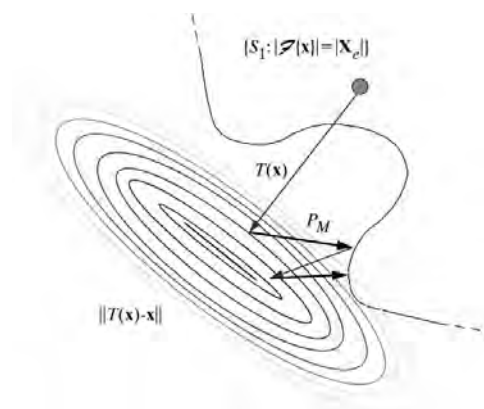

Fig. 3. Contours of $||T(\mathbf{x}) - \mathbf{x}||$ (assumed here to be locally convex) and the nonconvex set defined by the known constraint $|\mathbf{X}| = |\mathbf{X}_e|$ where $|\mathbf{X}_e|$ are the experimentally measured moduli.

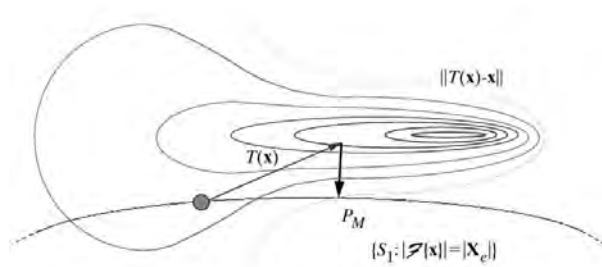

Fig. 4. Contours of FOM [*e.g.* $||T(\mathbf{x}) - \mathbf{x}||$, assumed here to be nonconvex] and a given set. The figure demonstrates how one iteration $P_M(T(\mathbf{x}))$ may result in an increase of the calculated FOM.

Since we know that the scattering comes from atoms, we have certain constraints on the possible values of $\mathbf{X}$, which are more conveniently written in real space in terms of $\mathbf{x}$ and nonlinear operators. For instance, we have the Sayre equation (Sayre, 1952), *i.e.*

$$T(\mathbf{x}) = C\mathbf{x}^2, \tag{18}$$

where C is a calculable constant or function that depends upon the number and type of atoms in a standard fashion. The fixed points of (18) are the set of all electron density maps $\mathbf{x}$ containing only the values 0 and $1/C$. Let P_M correspond to projecting a given set of values onto the measured $|\mathbf{X}|$, as illustrated in Fig. 5. The iterative sequence

$$x_{n+1} = P_M(T(x_n)) \tag{19}$$

corresponds to a 'generalized direct methods' algorithm and is clearly very similar to an iterative feasible set solution method. The application of an operator T represents a step closer to a fixed point of T (presuming T is nonexpansive; see below). By defining the functional $g(\mathbf{X})$ such that fixed points of T are minima of $g(\mathbf{X})$, equation (19) provides an iterative scheme for minimizing $g(\mathbf{X})$ consistent with the measured $|\mathbf{X}|$.

In a similar sense, we can describe structure completion in terms of some mapping M_A that projects onto the set of atomic coordinates d, *i.e.*

$$d_n = M_A(x_n). \tag{20}$$

The general solution to the crystallographic phase problem combines in some fashion equations (19) and (20), which can be represented diagramatically as indicated in Fig. 6.

A requirement of the proposed strong parallels between direct methods and the feasible set problem is nonexpansivity of the operator T. However, as defined in (18), T is expansive (*i.e.* the opposite of contractive) for many $\mathbf{x}$; one can consider T as a 'sharpening operator' in the sense that it increases large values of $\mathbf{x}$ and suppresses small ones. To overcome this problem, we note that (assuming all $|\mathbf{X}|$ are known)

$$P_M(T(x_n)) = P_M(\alpha T(x_n)), \tag{21}$$

where α is an arbitrary constant. Let us choose the renormalization constant α such that it minimizes some metric (not necessarily $||\ldots||$) of $\alpha T(\mathbf{x}) - \mathbf{x}$. The same metric can be interpreted as a log-likelihood or probability functional $g(\mathbf{x})$; it is minimal (zero) for $\mathbf{x}$ a fixed point of T, and we can associate with the values of this functional below some scalar value a set, the set S_0 discussed earlier. Use of a scaling by α can be considered as making the operator $\alpha T(\mathbf{x})$ 'conservative' in the sense that it conserves either the mean or the standard deviation, *e.g.*

$$\alpha = \langle x_n \rangle / \langle T(x_n) \rangle, \tag{22}$$

or α is taken to minimize either the L_1 or L_2 mean, *i.e.* minimize

$$L_p = \left[\sum |x_n - \alpha T(x_n)|^p \right]^{1/p} \tag{23}$$

or the equivalent form in reciprocal space taken over the measured reflections. Proving that this is nonexpansive for general direct-methods operators in a formal mathematical sense is not an easy task. Pragmatically, it can be tested numerically. For the simple Sayre-type operator using the mean from equation (22) does not give a nonexpansive operator, but using either an L_1 or L_2 mean minimization does. For later reference, the 'minimum relative entropy' (Cover & Thomas, 1991; Marks & Landree, 1998) or Kullback–Leibler distance (Kullback & Leibler, 1951) operator given by

$$T(\mathbf{x}) = \mathbf{x} \ln \mathbf{x} / \langle \mathbf{x} \rangle + \langle \mathbf{x} \rangle \tag{24}$$

is nonexpansive for any of the above renormalizations. Hereinafter, we will assume that the operators are nonexpansive within some range of $\mathbf{x}$ which can be enforced by renormalization, with the caveat that some

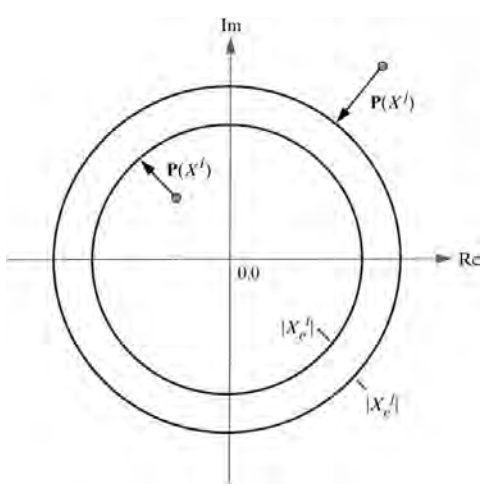

Fig. 5. Projection of $\mathbf{X}$ onto $|\mathbf{X}_e|$ in reciprocal space for two representative reflections X^i and X^j.

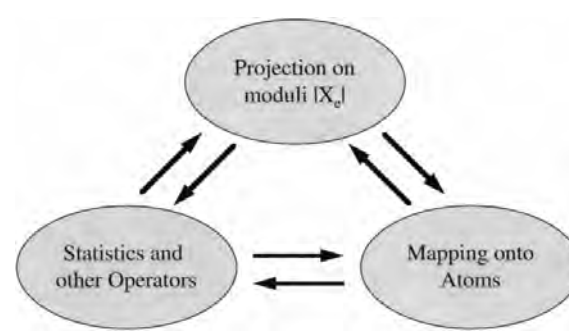

Fig. 6. Schematic representation of the relation between the different components necessary for structure determination.

more formal analysis would be very useful. Since there are many different fixed points of T for different atomic arrangements, the set of such fixed points is in general nonconvex, as is the log-likelihood functional set S_0. The feasible set may be nonconvex and discontinuous; see for instance Fig. 7.

One additional slightly complicated constraint needs to be introduced before we finalize our representation of direct methods. In addition to the statistical operators $T(\mathbf{x})$, we also have information about the shape of the distribution in reciprocal space. In a conventional Σ_2 sense, two large normalized (or unitary) structure factors are more likely to predict the phase of a third structure factor. If several Σ_2 relationships (Karle & Hauptman, 1956; Shmueli & Weiss, 1995) all predict approximately the same phase, there is a higher probability that this phase is correct and should be accepted for further steps in a phase extension. We need to encode this 'phase-extension constraint' in a more formal mathematical way.

One method is to consider the individual i components of the structure factors after each cycle, $\mathbf{X}^i_{n+1}$, relative to the experimental moduli $|\mathbf{X}^i_e|$. If the predicted amplitude is large, *i.e.* comparable to $|\mathbf{X}^i_e|$, this corresponds to the case in which the individual phase predictions for beam i are in good agreement and constructively interfere If this is the case, it is more probable that there is a correct prediction of a new phase. In a more formal way, we can specify the condition for accepting a new phase as

$$\alpha|\mathbf{X}^i_{n+1}| > \gamma_{n+1}|\mathbf{X}^i_e|, \quad 0 < \gamma_{n+1} < 1, \tag{25}$$

with α the scaling constant from before and γ some adjustable scalar. Introduction of this phase-extension constraint provides a connection between direct-methods approaches based on real-space operators, such as the Sayre method or the minimum relative-entropy operator, and reciprocal-space methods employing classical phase extension in an approach which gradually adds beams to the known set. (Note that exploiting the prior information available in the distribution in principle goes beyond simple phase extension.) The set of complex structure factors that obey equation (25) is not convex.

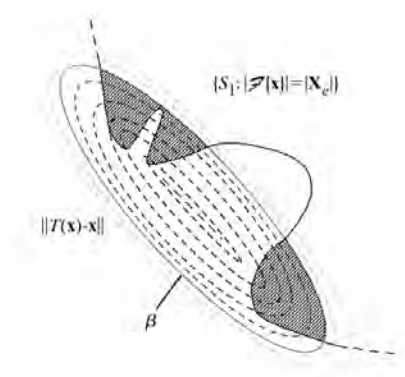

Fig. 7. Schematic illustration of the nonconvexity and discontinuous nature of the problem. Shown is the intersection for the nonconvex set (*e.g.* $|\mathbf{X}| = |\mathbf{X}_e|$) and the set defined by $||T(\mathbf{x}) - \mathbf{x}|| \leq \beta$ (assumed to be locally convex).

We can now state that the generalized (with renormalization) direct-methods algorithm is formally equivalent to an iterative feasible set method using a statistical or atomistic operator interpretable *via* a log-likelihood as a functional (hopefully, but not guaranteed to be nonexpansive) and projection onto the nonconvex set of experimental structure factors coupled with a phase-extension constraint. If we take $\mathbf{X}$ to be the structure factors, we have the original Sayre method (Sayre, 1952); if we take $\mathbf{X}$ to be the normalized structure factors, we have the tangent formula (Karle & Hauptman, 1956).

Some additional points merit mention. First, above we have assumed that the 'true' solution for the phases is a fixed point of the operator T, but in reality the operator is probabilistic in character. However, since we always have a downhill step (or at least not an uphill one) we will have a general error-reducing behavior. If the operator T moves all the way to a local minimum, this will always be the case. However, if instead it moves part way towards a minimum, the method may move slightly uphill after finding the best point depending upon exactly when the FOM (metric) is evaluated; see Fig. 4. This is not really a concern since we only need to find approximately correct values in general. Since around any local minimum we have a locally convex set, we will have strong convergence to that minimum assuming that it satisfies the other constraints.

Second, we note that the general FOM should be (as noted previously) an appropriate metric of the distance from all the constraints, with zero for an ideal fit. Since in reality there are measurement errors and errors in the scaling to give unitary or normalized structure factors, this will not be zero in the general case. There is no reason why this metric should be the classic Hilbert space norm defined above in equation (3), and in fact there are good reasons why it should not be. The Hilbert space norm is similar to a classical χ^2 refinement in that it is not tolerant of a few points with large errors (outliers). Other metrics, for instance the L_1 metric [equation (23)], are more robust in that they tolerate outliers more, similar to the classic R factor used in the refinement of atomic positions. (It is in this sense that the term Banach space, where norms other than $||\ldots||$ can be applied, may be better.)

Finally, the success of the method will depend upon how good is our choice of the operator T. To strengthen this, one can introduce a 'window function' $W(k)$ in reciprocal space (Marks & Landree, 1998), such that for (say) the unitary structure factors $U(k)$

$$U'(k) = W(k)U(k) \tag{26}$$

and

$$W(k) = \gamma \mathcal{F}^{-1} T(\mathcal{F} W(k)) \tag{27}$$

with $\mathcal{F}$ the Fourier transform operation and γ a constant. This enforces atomicity in the sense that it makes a set of atoms a fixed point of the operator. [For completeness, $W(k)$ need not exist for all possible operators, but only needs to be achievable within the limits of measurement and other errors, *e.g.* errors in the conversion from $F(k)$ to $U(k)$ or $E(k)$.]

Quite a few additional points about how the algorithm can be employed are worth mentioning.

(i) The simple algorithm of (19) does not exploit the relaxation parameter λ used in the more general statement of (17). Since it is exceedingly well documented in the literature that values of $\lambda > 1$ converge much better (Levi & Stark, 1987), it immediately follows that, for instance, the simple tangent formula should be over-relaxed to improve convergence. (Unfortunately, the literature is also clear that the optimum value of λ is dependent upon the exact problem being solved, and tends to be somewhat empirical in character.)

(ii) Rather than just using one operator, more than one operator can be employed. As an example, what is called the parallel projection method (PLC, pp. 206–209) combines a number of different operators $T_i(\mathbf{X})$ *via*

$$T(\mathbf{X}) = \sum w_i T_i(\mathbf{X}), \quad \sum w_i = 1. \tag{28}$$

In practice, we have found that combining the relative entropy operator and positivity gives better results than either alone (see below).

(iii) Both structure completion methods and algorithms such as *Shake-and-Bake* (Miller *et al.*, 1994), by which atoms are introduced as elements of the structure, incorporate convex constraints into the problem. For instance, the set of all $\mathbf{x}$ with atoms at certain locations is a convex set, and as the number of known atoms increases, this set shrinks in size.

(iv) One can create new algorithms using known convex set constraints. For instance, we may know (for instance from low-resolution phasing or electron microscopy) that there are no atoms at certain locations. This is similar to what is called a 'support constraint' (Fienup, 1987; Hayes, 1987) in the image-processing literature, and is a convex set. Furthermore, it is known that in many cases (see Hayes, 1987) the phase problem is fully defined by such a support constraint plus knowledge of the structure-factor moduli. (Positivity can also be exploited.) If an upper bound to the charge density is known (*e.g. via* the heaviest atom), this can also be exploited, as can a variety of statistical projections if the measurement errors are known to be Gaussian (PLC, pp. 193–198).

4. Numerical examples

We will present here a few examples that will illustrate the application of the feasible set approach.

4.1. *Structure completion*

Suppose we know a subset of the atomic sites and apply the chemical constraint that there are no atoms within some region of each of the atoms. This is a convex constraint which can be written *via* the projection

$$P(\mathbf{x}) = \begin{cases} \rho_k & \text{for } \mathbf{x} \in S_k \\ \mathbf{x} & \text{otherwise,} \end{cases} \tag{29}$$

where ρ_k are the known values and S_k the set of points within a certain radius of each atomic position. Coupling this with the positivity P_P,

$$P_P(\mathbf{x}) = \begin{cases} \mathbf{x} & \text{for } \mathbf{x} > 0 \\ 0 & \text{for } \mathbf{x} < 0, \end{cases} \tag{30}$$

gives a combined projection

$$P(\mathbf{x}(r)) = \begin{cases} \rho_k & \text{for } \mathbf{x} \in S_k \\ \mathbf{x} & \text{for } \mathbf{x} \notin S_k,\ \mathbf{x} > 0 \\ 0 & \text{for } \mathbf{x} \notin S_k,\ \mathbf{x} < 0. \end{cases} \tag{31}$$

Finally, adding in projection onto the moduli, *i.e.*

$$P_M(\mathbf{X}(k)) = \mathbf{X}(k)|\mathbf{X}_e(k)|/|\mathbf{X}(k)|, \tag{32}$$

where $|\mathbf{X}_e(k)|$ are the experimental moduli, gives an algorithm very similar to the Gerchberg–Saxton (Gerchberg & Saxton, 1972; Gerchberg, 1974) and Fienup algorithm (Fienup, 1978, 1987).

To illustrate application of this algorithm, Fig. 8 shows results for perbromo-, perchloro- and perfluorophthalocyanine complexes with copper with kinematical electron diffraction data and an over-relaxation parameter of $\lambda = 1.95$, assuming that only the halogen atoms are known and with *c2mm* symmetry imposed (no origin definition or other phases were set). Using the deviation from the projection of equation (31) as a FOM, the best (FOM) results for 30 different trials, each ten cycles long with initial random values, are shown. (The moduli projection is not convex, so local minima will be found.) Almost independent of the starting point, results similar to Fig. 8(*a*) were obtained for the bromine complex and are close to exact restorations of the true $\mathbf{x}$ with well defined atomic features. For the chlorine derivative, several of the results were similar to Fig. 8(*b*) and again approach restorations of $\mathbf{x}$. For the fluorine compound (Fig. 8*c*), the results are much worse but still relatively close, and would improve if a larger number of tests were performed. In none of these examples was atomicity in the unknown portions of $\mathbf{x}$ used.

This is a very simple algorithm, similar in many respects to approaches that have previously been used in the literature (*e.g.* Woolfson & Fan, 1995; Millane, 1996; Millane & Stroud, 1997). The relevant point here is that

the definition in terms of a feasible set approach is an illustration of the link between this and already established structure completion methods.

4.2. *Restoration of unmeasured reflections*

A particular problem with surface diffraction data (Marks *et al.*, 1997, 1998; Gilmore *et al.*, 1997; Collazo-Davila *et al.*, 1997, 1998) is that some of the strongest reflections may not have been measured; without these included in some sense in the calculation it may not be possible to solve the structure. If we know (or can link statistically) the phases of the measured reflections, we have a good estimation of the phases of the unmeasured reflections. For a given set of estimated phases, the set of all moduli for the unmeasured reflections is a convex set. Coupling this with positivity, we have the intersection of two convex sets for the unmeasured reflections, so the problem is in principle well defined.

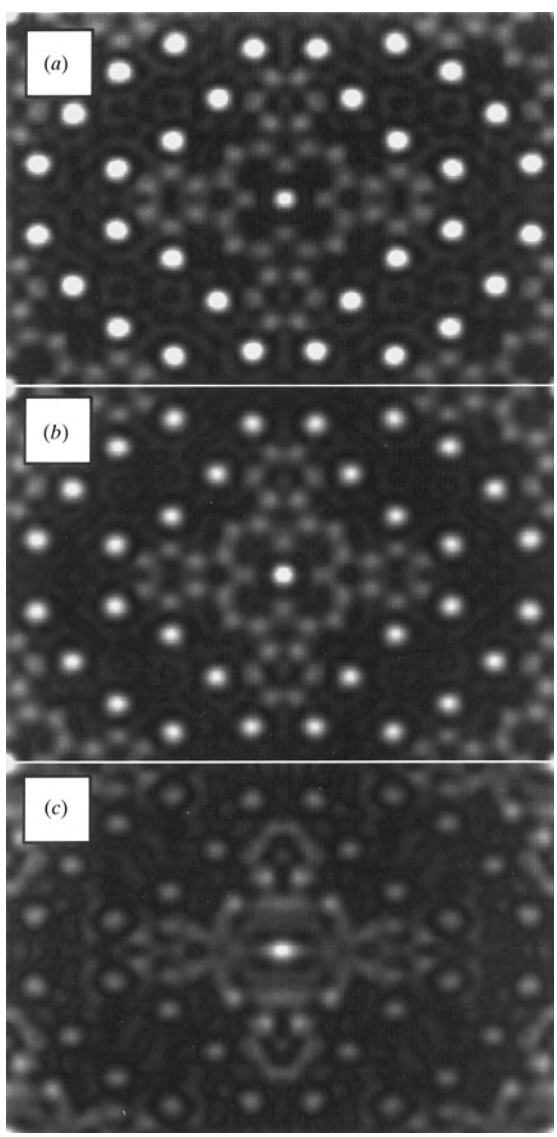

Fig. 8. Completed structure density maps of (*a*) perbromophthalocyanine, (*b*) perchlorophthalocyanine and (*c*) perfluorophthalocyanine.

Table 1. *Comparison of the FOM and number of iterations necessary to converge to the minimum solution for the minimum relative-entropy algorithm and the hybrid algorithm with positivity constraint*

	Minimum entropy only		Hybrid algorithm	
λ	FOM	No. of iterations†	FOM	No. of iterations†
0.6	0.4351	8	0.4392	5
0.7	0.4219	7	0.4392	5
0.8	0.3952	12	0.4382	5
0.9	0.3761	9	0.4340	6
1.0	0.3736	9	0.4299	6
1.1	0.3390	12	0.4269	7
1.2	0.3358	10	0.4202	6
1.3	0.3365	10	0.4054	11
1.4	0.3423	9	0.3943	11
1.5	0.3503	6	0.3791	7
1.6	0.3466	7	0.3530	10
1.7	0.3452	6	0.3514	10
1.8	0.3498	6	0.3163	15
1.9	0.3695	6	0.3161	11
2.0	0.3752	8	0.4273	7
2.1	0.3779	7	0.4273	7

† No. of iterations refers to the number of iterations for the algorithm to converge to its minimum value.

The central issue is how to scale the unmeasured moduli, and here the scaling term α used previously works very well in practice. The full algorithm can be written as

$$T(\mathbf{x}) = \begin{cases} \mathbf{x}\ln[\mathbf{x}/\langle\mathbf{x}\rangle] + \langle\mathbf{x}\rangle & \text{for } \mathbf{x} > 0 \\ \langle\mathbf{x}\rangle & \text{otherwise,} \end{cases} \tag{33}$$

$$X_{n+1} = \begin{cases} \mathbf{Q}_n|\mathbf{X}_e|/|\mathbf{Q}_n| & \text{for } \mathbf{X} \in S_M, \\ & \quad |\mathbf{X}_{n+1}| > (\gamma_n/\alpha)|\mathbf{X}_e^i| \\ \alpha\mathcal{F}^{-1}T(\mathbf{x}_n) & \text{for } \mathbf{X} \notin S_M,\ (h,k,l) \in D \\ 0 & \text{for } \mathbf{X} \notin S_M,\ (h,k,l) \notin D \\ |\mathbf{X}|\exp(i\boldsymbol{\varphi}_F) & \text{for } \mathbf{X} \in S_F, \end{cases} \tag{34}$$

with

$$\mathbf{Q}_n = \mathbf{X}_n + \lambda[P_M(\alpha\mathcal{F}^{-1}T(x_n)) - X_n], \tag{35}$$

where S_F is the set of fixed (*e.g.* origin defining) reflections with phases $\boldsymbol{\varphi}_F$, S_M is the set of measured reflections and D the set of all possible reflections (hkl) which lie within an elliptical aperture in two-dimensional space, or an ellipsoidal aperture in three-dimensional space, that includes the measured reflections. The FOM used is a normalized L_1 mean taken over the set $S_M \cap S_F$, *i.e.*

$$\text{FOM} = \left(\sum|\mathbf{X}_n - \alpha\mathbf{X}_{n+1}|\right)\Big/\sum|\mathbf{X}_n|, \tag{36}$$

and the constant α minimizes (35). A phase-extension constraint as in (25) was used with

$$\gamma_n = 0.3\exp(-n/2). \quad (37)$$

For reference, the combination of (34) and (35) includes over-relaxation of the phases as illustrated in Fig. 9.

As an example, Figs. 10 and 11 show the results of a phase extension with experimental electron diffraction data for the Si (111) 7×7 surface (Gilmore *et al.*, 1997) using the minimum relative entropy operator with initial phases to about 2.8 Å and phase extension to approximately 0.5 Å resolution. Comparison of the maps without and with projected values for the unmeasured reflections (Figs. 10*a* and 10*b*, respectively) shows their importance; the unmeasured (7,7) reflections have structure factors 3–5 times larger than any of the other reflections and without them not all the atoms are visible (arrowed in Fig. 11). For reference, an over-relaxation of $\lambda = 1.2$ was used and Figs. 10(*c*) and 11(*c*) are for a different algorithm, discussed below (§4.4).

4.3. *Effect of over-relaxation*

To demonstrate the role of over-relaxation, Table 1 shows the total number of iterations to convergence of a calculation as a function of the relaxation parameter [*e.g.* equation (17)] for the same data as in Figs. 10 and 11, and also the final FOM. In all cases, the FOM (metric) is the L_1 mean [equation (36)] appropriately scaled in reciprocal space over the measured reflections. Here over-relaxation was only used for the phases (Fig. 9). [While the result shown here is for a simple centrosymmetric case, similar results have been found for noncentrosymmetric structures and by *ab initio* approaches using a genetic algorithm (Landree *et al.*, 1997) to control the initial phase choices.] Interestingly, the 'better' values for the relaxation parameter tend to give slightly smaller FOMs – a reproducible effect that we have seen in many cases.

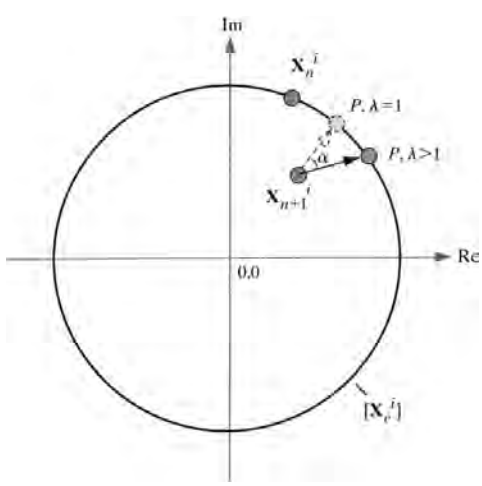

Fig. 9. Schematic representation of over-relaxation of the phase for X^i, shown for $\lambda > 1$, $\alpha > 0$ (*e.g.* $\lambda = 1$, $\alpha = 0$).

4.4. *Hybrid relative entropy/positivity algorithm*

As an example of combining different projections, consider the relative entropy operator [equation (33)] and the positivity projection [equation (30)]. While positivity does not have to lead to atomistic solutions, it is still approximately correct. (Since experimental data are band-width limited, equivalent to superimposing an aperture in reciprocal space, the charge density for a single atom with the measured reflections is not everywhere positive.) Using a combined operator as in

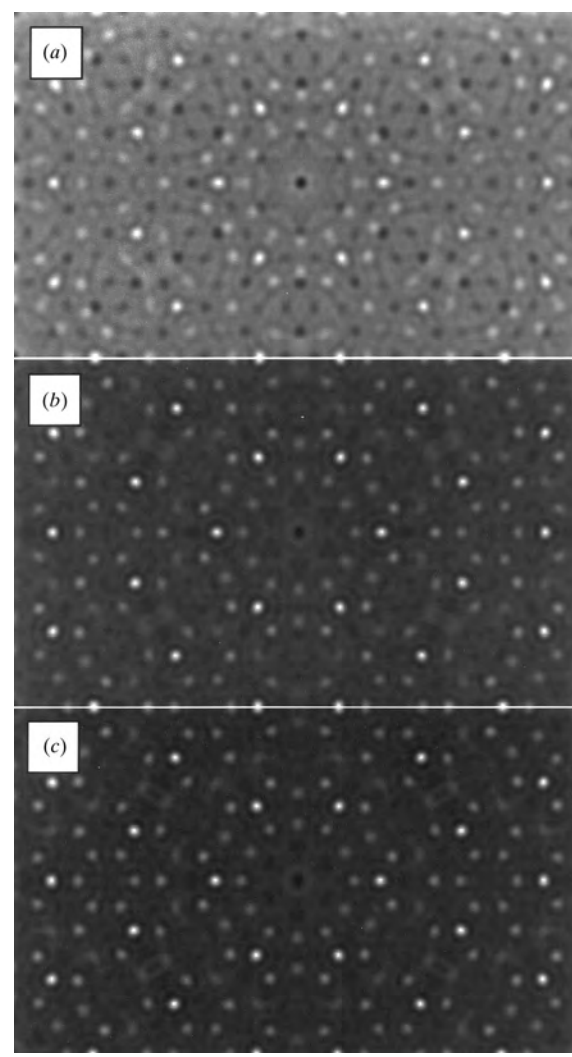

Fig. 10. Minimum relative-entropy phase-extension maps of the Si 7×7 (*a*) without and (*b*) with projected values for the unmeasured reflections, and (*c*) for a hybrid minimum relative-entropy algorithm with projected values for the unmeasured reflections and positivity constraint.

equation (28), the weighting scheme for the entropy (w_e) and positivity (w_p),

$$w_e = \max(\alpha, 1),$$
$$w_p = 1 - w_e, \tag{38}$$

with α the renormalization term from before, is, empirically, effective. Figs. 10(*c*) and 11(*c*) compare calculations with this hybrid algorithm and without the positivity (Figs. 10*b* and 11*b*) for a relaxation of 1.8, and the final FOM is shown *versus* relaxation parameter in Table 1. In this case, the effects are relatively subtle and are a suppression of background noise improving the contrast (peak to background ratio); we have seen similar and more substantive effects in other calculations with different structures.

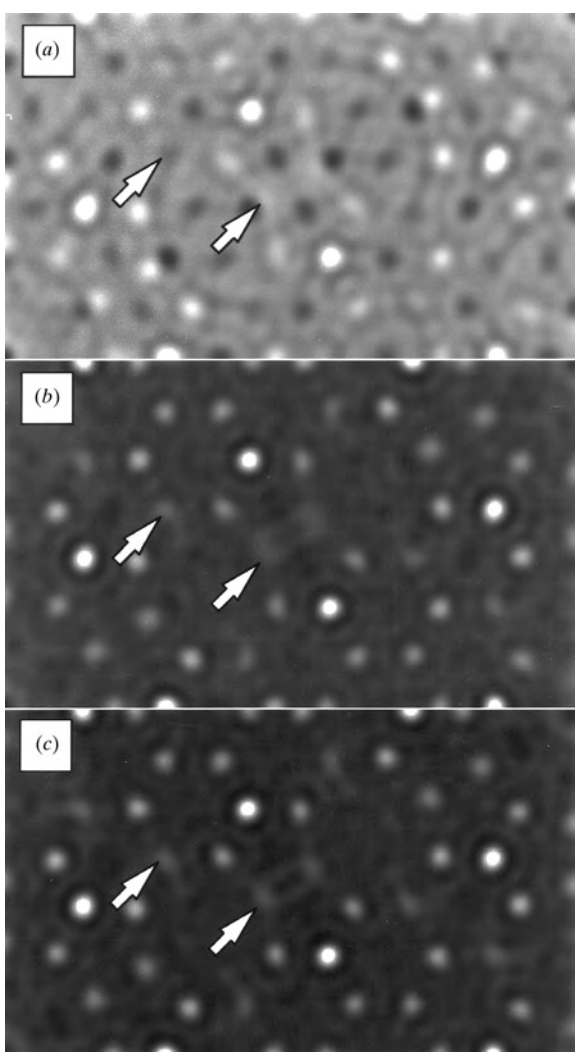

Fig. 11. Magnified regions from Figs. 10(*a*)–(*c*). Atoms previously unseen due to the exclusion of the unmeasured reflections are arrowed for reference.

5. Discussion

This paper is a start towards combining the feasible set approach and direct methods, not the end. The connection is clearly there and enables both more formal mathematical analysis of the algorithms as well as tools (*e.g.* over-relaxation) for the optimization and design of new algorithms. In no way does this approach disagree with or dispute any of the existing statistical relationships available in the literature; instead it investigates the actual algorithm wherein they are used.

Many topics remain for further work, both in terms of the formal mathematics and in terms of implementation. To discuss the mathematics first, clearly the issue of nonexpansiveness of the renormalized operators needs to be analyzed in more detail. While the 'proof', *i.e.* numerical tests, given herein is relatively weak, one does of course have a stronger empirical proof, *i.e.* direct methods have been known to work for more than thirty years. Certainly the problem of convergence (and other characteristics) for nonconvex sets needs to be addressed in more detail than it appears to have received to date. Indeed, several papers consider the nonconvex problem to be prohibitively expensive in computer time and therefore unfeasible. The multi-solution approaches developed over the years for direct methods, such as magic numbers (White & Woolfson, 1975), error-correcting codes (Gilmore & Nicholson, 1994), simulated annealing (Sheldrick, 1990; Bhat, 1990) and genetic algorithms (Landree *et al.*, 1997), contradict this view. By its very character it may never be possible to recast the crystallographic phase problem in terms of convex sets which have a single solution. However, we do not consider this to be an important issue since global search algorithms are now relatively mature and not prohibitively expensive in computer time.

Much work could usefully be performed converting some of the existing direct-methods FOMs to an operator (or functional) format for incorporation, assuming that they are nonexpansive. Rather than a multiple-FOM approach as currently used, a multiple operator/projection approach could be employed similar to the hybrid method described above. While this does increase the programming complexity, our experience to date is that the convergence rate and faithfulness of the solutions are worth the effort. Furthermore, such an approach is intrinsically parallel, so may be able to exploit developments in computer architecture rather well.

There are many possible new algorithms that can be constructed, not all of which will work. Within the feasible set approach, one can formulate methods for including error estimates for the measurements. For instance, instead of projecting onto the measured structure factors ($|\mathbf{X}_e|$), one could use the projection

$$P(\mathbf{X}) = \mathbf{Y} = \mathbf{X}/|\mathbf{X}|[\beta|\mathbf{X}_e| + (1-\beta)|\mathbf{X}|], \tag{39}$$

where β is chosen such that

$$\chi^p = 1 = 1/M \sum(|\mathbf{X}_e| - |\mathbf{Y}|)^p/\sigma^p \qquad (40)$$

with σ the measurement error, M the number of points and $p = 1$ or 2 for robust or Gaussian errors. Following this projection, the moduli used for the next cycle will lie distributed about the measured values in a fashion consistent with the experimental data.

Methods for accelerating convergence may also be relevant. For instance, we use Fourier transforms and real-space operators; the Fourier transforms constitute the rate-limiting step. Using two steps of the atomistic operators prior to back transforming might substantially improve performance.

One can also formulate methods for cases in which there are overlapping reflections, for instance twinning or for surfaces when reflections from the surface layer coincide with bulk reflections. As an example of the latter, the scattering from the bulk can be calculated and combined with the estimated coincident surface complex value and projected onto the measured coherent addition of the two; in numerical tests that will be discussed more elsewhere, this works rather well. Going one step further, there is no need to insist upon strict conformance with kinematical scattering and extensions are possible for dynamical electron diffraction. Provided that appropriate operators and/or functionals can be defined, solutions of the dynamical phase problem are possible using modified direct methods, as has already been shown for well resolved zone-axis orientations (Sinkler *et al.*, 1998; Sinkler & Marks, 1999).

Of course, not all methods suggested by the feasible set approach will work; there are always questions of stability which are not easy to analyze in a formal mathematical sense. However, it is the belief of the authors that there are enough new avenues that some will outperform, perhaps rather substantially, existing algorithms.

The authors would like to thank P. L. Combettes for particularly useful information, and also thank J. Fienup, G. Gasper, A. K. Rajagopal and J. C. H. Spence for their comments. This work was supported by the National Science Foundation under grant numbers DMR-9705081 (LDM and EL) and DMR 91–20000 (WS).

References

Bhat, N. T. (1990). *Acta Cryst.* A**46**, 735–742.
Bricogne, G. (1984). *Acta Cryst.* A**40**, 410–445.
Chrétien, S. & Bondon, P. (1996). *Num. Func. Anal. Optim.* **17**, 37–56.
Collazo-Davila, C., Grozea, D. & Marks, L. D. (1998). *Phys. Rev. Lett.* **80,** 1678–1681.
Collazo-Davila, C., Marks, L. D., Nishii, K. & Tanishiro, Y. (1997). *Surf. Rev. Lett.* **4**, 65–70.
Combettes, P. L. (1996). *Adv. Imag. Elec. Phys.* **95**, 155–270.
Combettes, P. L. & Trussell, H. J. (1990). *J. Optim. Theory Appl.* **67**, 487–507.
Cover, T. M. & Thomas, J. A. (1991). *Elements of Information Theory*. New York: John Wiley.
Dainty, J. C. & Fienup, J. R. (1987). In *Image Recovery: Theory and Application*, edited by H. Stark. Orlando: Academic Press.
Dorset, D. L. (1996). *Acta Cryst.* A**52**, 753–796.
Dynes, R. C. & Fulton, T. A. (1971). *Phys. Rev. B*, **3**, 3015–3023.
Fienup, J. R. (1978). *Opt. Lett.* **3**, 27–29.
Fienup, J. R. (1987). *J. Opt. Soc. Am.* **4**, 118–123.
Gerchberg, R. W. (1974). *Opt. Acta*, **21** 709–720.
Gerchberg, R. W. & Saxton, W. O. (1972). *Optik* (*Stuttgart*), **35**, 237–246.
Giacovazzo, G. (1980). *Direct Methods in Crystallography*. New York: Plenum Press.
Gilmore, C. J. (1996). *Acta Cryst.* A**52**, 561–589.
Gilmore, C. J., Marks, L. D., Grozea, D., Collazo-Davila, C., Landree, E. & Twesten, R. D. (1997). *Surf. Sci.* **381**, 77–91.
Gilmore, C. J. & Nicholson, W. V. (1994). *Trans. Am. Crystallogr. Assoc.* **30**, 15.
Gubin, L. G., Polyak, B. T. & Raik, E. V. (1967). *USSR Comput. Math. Math. Phys.* **7**, 1–24.
Hayes, M. H. (1987). *Image Recovery, Theory and Application*, edited by H. Stark, pp. 195–230. Orlando: Academic Press.
Herman, G. T. (1980). *Image Reconstruction from Projections – the Fundamentals of Computerized Tomography*. New York: Academic Press.
Kaczmarz, S. (1937). *Bull. Acad. Sci. Pol.*, A**35**, 355–357.
Karle, J. & Hauptman, H. (1956). *Acta Cryst.* **9**, 635–651.
Kullback, S. & Leibler, R. A. (1951). *Ann. Math. Stat.* **22**, 79–86.
Landree, E., Collazo-Davila, C. & Marks, L. D. (1997). *Acta Cryst.* B**53**, 916–922.
Levi, A. & Stark, H. (1987). *Image Recovery, Theory and Application*, edited by H. Stark, pp. 277–320. Orlando: Academic Press.
Marks, L. D., Bengu, E., Collazo-Davila, C., Grozea, D., Landree, E., Leslie, C. & Sinkler, W. (1998). *Surf. Rev. Lett.* **5**, 1087–1106.
Marks, L. D. & Landree, E. (1998). *Acta Cryst.* A**54**, 296–305.
Marks, L. D., Plass, R. & Dorset, D. L. (1997). *Surf. Rev. Lett.* **4**, 1–8.
Millane, R. P. (1996). *J. Opt. Soc. Am.* A**13**, 725–734.
Millane, R. P. & Stroud, W. J. (1997). *J. Opt. Soc. Am.* A**14**, 568–579.
Miller, R., Gallo, S. M., Khalak, H. G. & Weeks, C. M. (1994). *J. Appl. Cryst.* **27**, 613–621.
Ramkrishna, D. & Amundson, N. R. (1985). *Linear Operator Methods in Chemical Engineering*, p. 5. New Jersey: Prentice Hall.
Sayre, D. (1952). *Acta Cryst.* **5**, 60–65.
Sezan, M. I. (1992). *Ultramicroscopy*, **40**, 55–67.
Sheldrick, G. M. (1990). *Acta Cryst.* A**46**, 467–473.
Shmueli, U. & Weiss, G. H. (1995). *Introduction to Crystallographic Statistics.* Oxford University Press.
Sinkler, W., Bengu, E. & Marks, L. D. (1998). *Acta Cryst.* A**54**, 591–605.
Sinkler, W. & Marks, L. D. (1999). *Ultramicroscopy*, **75**, 251–268.

Stark, H. (1987). *Image Recovery: Theory and Applications.* Orlando: Academic Press, Inc.
White, P. S. & Woolfson, M. M. (1975). *Acta Cryst.* A**31**, 53–56.
Woolfson, M. M. (1987). *Acta Cryst.* A**43**, 593–612.
Woolfson, M. M. & Fan, H. (1995). *Physical and Non-Physical Methods of Solving Crystal Structures.* Cambridge University Press.
Youla, D. C. (1987). *Image Recovery, Theory and Application*, edited by H. Stark, pp. 29–77. Orlando: Academic Press.