INSTITUTE OF PHYSICS PUBLISHING INVERSE PROBLEMS

Inverse Problems **21** (2005) 37–50 doi:10.1088/0266-5611/21/1/004

Relaxed averaged alternating reflections for diffraction imaging

D Russell Luke

Department of Mathematical Sciences, University of Delaware, Newark, DE 19716-2553, USA

E-mail: rluke@math.udel.edu

Received 11 May 2004, in final form 8 September 2004
Published 9 November 2004
Online at stacks.iop.org/IP/21/37

Abstract

We report on progress in algorithms for iterative phase retrieval. The theory of convex optimization is used to develop and to gain insight into counterparts for the nonconvex problem of phase retrieval. We propose a relaxation of averaged alternating reflectors and determine the fixed-point set of the related operator in the convex case. A numerical study supports our theoretical observations and demonstrates the effectiveness of the algorithm compared to the current state of the art.

1. Introduction

The phase retrieval problem is a classical inverse problem in optics that has received renewed interest in applications to nonperiodic scatterers and macromolecules. In a typical x-ray crystallography experiment, for example, a crystalline specimen is illuminated with a monochromatic x-ray and the resulting diffraction pattern is recorded. In the far field of the crystal, that is, in the region of validity of the Fraunhofer approximation, the complex amplitude of the diffracted x-rays is equal to the (scaled) Fourier transform of the electron density distribution of the specimen. The problem is that only the intensity of the diffracted field can be measured. The missing phase information is critical for determining the electron density. In some cases, such as x-ray crystallography of 'small' (relative to the source wavelength) periodic molecules, it is possible to determine the electron density by what are referred to as *direct methods* [12]. For large macromolecules and nonperiodic structures, however, one must rely on numerical techniques for reconstructing the missing phase. So-called *iterative transform methods* pioneered by Gerchberg and Saxton [11] and Fienup [10] are well-established generic iterative techniques for recovering the phase in a variety of settings. Recent developments in imaging [9, 13, 14, 19–22] have placed a premium on improving the efficiency and stability of these types of algorithms. This is the principal motivation of our work. For a derivation of the phase problem from first principles, the reader

 © 2005 IOP Publishing Ltd Printed in the UK

is referred to [15] and references therein. For a review of the phase problem in crystallography, see [16].

In this work we derive a stable and fast new strategy for phase retrieval, that we call relaxed averaged alternating reflection (RAAR), that falls under the category of iterative transform methods [15]. The motivation for the RAAR algorithm comes from recent work in which another new algorithm, the hybrid projection reflection algorithm (HPR), was presented [4]. The HPR algorithm was originally conceived as a single-parameter relaxation of the well-known Douglas–Rachford algorithm applied to phase retrieval. The HPR algorithm can also be viewed a special case of the three-parameter *difference map* recently proposed by Elser [7].

There are two fundamental and distinct issues that accompany these algorithms. The first is the incorporation of *a priori* information into the constraint structure of the algorithms. The second is the choice of algorithm parameter values. Regarding the first issue, it is difficult to overestimate the effect of the constraints on the mathematical properties and performance of the algorithms. There have been several studies on the choice of constraints in applications to crystallography [8, 17, 18]. This issue is complicated by the algorithmic formulation that is conventional in the optics community, where physical rationale serves as the principal motivation and guide for algorithm design. We use a simple example to illustrate how seemingly minor changes in the physical domain constraints can lead to algorithms that appear very different when written in the conventional optical format. This has caused some confusion in the literature which we hope to clarify through an examination of the abstract algorithmic structures behind the leading techniques. The choice of parameters also has a dramatic impact on the mathematical properties of the algorithms and hence performance. Physical insight often provides the best (and only) basis for choosing values for the algorithm parameters, but this is not always available or reliable. In the case of the HPR algorithm, our numerical experiments have not provided an empirical basis upon which to make recommendations. A more mathematically rigorous approach also appears to be difficult and has been found in only a few very special cases. For instance, in the convex setting the convergence properties of the HPR algorithm are known for the unrelaxed case [5]. For the relaxed HPR algorithm and the more general difference map a complete and mathematically rigorous analysis has yet to be found. To circumvent the analytical barriers facing the difference map and the HPR algorithm, we introduce the RAAR algorithm, which is conceptually simple, analytically tractable and easy to implement; moreover, it outperforms the current state of the art. While the RAAR algorithm coincides with the HPR algorithm in a limiting case, it does not fall in the class of algorithms covered by Elser's difference map framework.

A precise statement of the leading algorithms is given in section 2. In this same section we provide a terse outline of the mathematical justification for the RAAR algorithm. In section 3 we demonstrate the effectiveness of the algorithm and make practical recommendations for implementation.

2. Phase retrieval and iterative transform algorithms

2.1. Phase retrieval

Returning to the crystallography example, we consider the problem of recovering the electron density of a crystal, denoted by u_*, from the diffraction pattern it produces upon illumination by a monochromatic x-ray source. In this setting it is natural to include some *a priori* assumptions, namely that u_* is a real-valued, nonnegative function supported on some prescribed bounded set D, that is $\mathcal{L} \ni u_* : \mathbb{Z}^N \to \mathbb{R}_+$ with $\operatorname{supp}(u_*) \subset D \subset \mathbb{Z}^N$. Here $\mathcal{L}$ is a Hilbert space of square integrable functions, $\mathbb{Z}^N$ is the domain—in this case the *physical domain*—corresponding to discrete (i.e. sampled) waves, $\mathbb{R}_+$ is the positive reals and $\mathrm{supp}(u_*)$ is the support of u_*. Writing this in terms of constraints, we have $u_* \in S_+ \subset \mathcal{L}$, where S_+ is the set of nonnegative functions in $\mathcal{L}$ with support on D. In many crystallographic settings the nonegativity of u_* is less important than the support. It is common, therefore, to require only that the functions be supported on D. In this case, we denote the corresponding constraint set by S. The sets S and S_+ are referred to as the *physical domain constraints*. The other constraint we consider comes from the data, m, which we presume consist of noisy magnitude measurements in the far field, thus m is proportional to the modulus of the Fourier transform of u_*. We therefore refer to the domain of the image data m as the *Fourier domain*. In terms of constraint sets, we write that $u_* \in M$ where $M = \{v \in \mathcal{L} \mid |\mathcal{F}v| = m\}$ and $\mathcal{F}v$ denotes the discrete Fourier transform of v. We shall refer to the set M as the *Fourier*, or *image domain constraint*. Note that S_+ is a *convex* set, while M is *nonconvex*. It is the nonconvexity of the magnitude constraint that does not allow us to transfer classical convergence results for the most common algorithms to the case of phase retrieval. For further discussion see [3].

2.2. *Iterative transform algorithms*

We formulate the problem of phase retrieval as a feasibility problem:

$$\text{find } u \in S_+ \cap M. \tag{1}$$

Iterative transform techniques are built upon combining projections onto the sets S_+ and M in some fashion. While they are seldom written as fixed-point algorithms, iterative transform algorithms can usually be put into the form $u_{n+1} = \mathcal{T} u_n$ where $\mathcal{T}$ is a generic operator in which the projections and averaging operations are embedded (see [3, 4]). For added control and flexibility, one often includes a *relaxation* strategy parametrized by β. We write the relaxed operator with generic, single parameter relaxation strategy $\mathcal{V}$ (there can be infinitely many such strategies) as $\mathcal{V}(\mathcal{T}, \beta)$. In order to effectively exploit relaxations for improved algorithm performance, it is necessary to understand the mathematical properties of the operator $\mathcal{V}(\mathcal{T}, \beta)$—first and foremost of these is the characterization of the set of fixed points, $\mathrm{Fix}\mathcal{V}(\mathcal{T}, \beta)$. We return to this issue at the end of this section.

The operators we study are built upon projectors and reflectors. Denote by P_C an arbitrary but fixed selection, or *projector*, from the possibly multi-valued *projection* onto a subset C of $\mathcal{L}$. Closely related is the corresponding *reflector* with respect to C

$$R_C = 2P_C - I,$$

where I is the identity operator. By definition, for every $u \in \mathcal{L}$, $P_C(u)$ is the midpoint between u and $R_C(u)$. Specializing to our application, the projector, $P_M u$, of a signal $u \in \mathcal{L}$ onto the Fourier magnitude constraint set M is given by

$$P_M(u) = \mathcal{F}^{-1}(\widehat{v}_0), \qquad \text{where} \qquad \widehat{v}_0(\xi) = \begin{cases} m(\xi)\dfrac{\mathcal{F}u(\xi)}{|\mathcal{F}u(\xi)|}, & \text{if } \mathcal{F}u(\xi) \neq 0; \\ m(\xi), & \text{otherwise.} \end{cases} \tag{2}$$

Here, $\mathcal{F}^{-1}$ is the discrete inverse Fourier transform and $\widehat{v}_0$ a selection from the multi-valued Fourier domain projection. For further discussion of this projector see Luke *et al* [15, corollary 4.3] and [6]. We return to the issue of multivaluedness of the magnitude projection in section 3. The projection of a signal $u \in \mathcal{L}$ onto S_+ is single-valued (since S_+ is convex), and is given by

$$(\forall x \in \mathbb{Z}^N) \qquad \big(P_{S_+}(u)\big)(x) = \begin{cases} \max\{0, u(x)\}, & \text{if } x \in D; \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

One of the best known iterative transform algorithms is Fienup's hybrid input–output algorithm (HIO) [10]. We use this as our benchmark for performance. In the present setting, HIO is given as

$$(\forall x \in \mathbb{Z}^N) \qquad u_{n+1}(x) = \begin{cases} (P_M(u_n))(x), & \text{if} \quad x \in D \quad \text{and} \\ & (P_M(u_n))(x) \geqslant 0; \\ u_n(x) - \beta_n (P_M(u_n))(x), & \text{otherwise.} \end{cases} \tag{4}$$

There have been several attempts to identify the HIO algorithm with a broader class of relaxation strategies that can be written as fixed-point iterations, that is, in the form $u_{n+1} = \mathcal{V}(\mathcal{T}, \beta_n)u_n$. Bauschke *et al* [3] proved that, when only a support constraint, as opposed to support and nonegativity, is applied in the physical domain, then the HIO algorithm with $\beta = 1$ corresponds to the classical Douglas–Rachford algorithm for which convergence results in the convex setting are well known. In a subsequent article Bauschke *et al* [4] proved that, for physical domain support constraints only, the HIO algorithm corresponds to a particular relaxation of the Douglas–Rachford algorithm, that is

$$(\forall x \in \mathbb{Z}^N) \qquad u_{n+1}(x) = \begin{cases} (P_M(u_n))(x), & \text{if } x \in D \\ u_n(x) - \beta_n (P_M(u_n))(x), & \text{otherwise,} \end{cases} \tag{5}$$

is equivalent to

$$u_{n+1} = \tfrac{1}{2}(R_S(R_M + (\beta_n - 1)P_M) + I + (1 - \beta_n)P_M)(u_n). \tag{6}$$

Independent of these results, Elser [7] demonstrated the correspondence between the HIO algorithm with only support constraints in the physical domain and the difference map,

$$u_{n+1} = (I + \beta(P_S((1 - \gamma_2)P_M - \gamma_2 I) + P_M((1 - \gamma_1)P_S - \gamma_1 I)))(u_n), \tag{7}$$

for the case where $\gamma_1 = -1$ and $\gamma_2 = 1/\beta$. It has been incorrectly assumed, however, that the correspondence between the difference map and the HIO algorithm carries over to the case of support and nonnegativity constraints—that is, upon replacing P_S in (7) with P_{S_+}, one obtains the HIO algorithm for nonnegativity constraints. Instead, this simple substitution of the constraints in the difference map with $\gamma_1 = -1$ and $\gamma_2 = 1/\beta$,

$$u_{n+1} = \left(I + \beta\left(P_{S_+}\left(\left(1 - \frac{1}{\beta}\right)P_M - \frac{1}{\beta}I\right) + P_M(2P_{S_+} + I)\right)\right)(u_n), \tag{8}$$

yields the hybrid projection reflection (HPR) algorithm proposed by Bauschke *et al* in [4]:

$$u_{n+1} = \tfrac{1}{2}\big(R_{S_+}(R_M + (\beta_n - 1)P_M) + I + (1 - \beta_n)P_M\big)(u_n). \tag{9}$$

In [4, proposition 2] it is shown that (9) is equivalent to

$$(\forall x \in \mathbb{Z}^N) \qquad u_{n+1}(x) = \begin{cases} (P_M(u_n))(x), & \text{if} \quad x \in D \quad \text{and} \quad (R_M(u_n))(x) \\ & \geqslant (1 - \beta_n)(P_M(u_n))(x); \\ u_n(x) - \beta_n (P_M(u_n))(x), & \text{otherwise.} \end{cases} \tag{10}$$

It is easy to see by comparing (10) with (4) that these are fundamentally different algorithms. Moreover, a reformulation of (4) in terms of a fixed-point iteration does not appear to be possible due to the nonlinearity of the P_{S_+} operator [4].

Algorithmic prescriptions of the form (4), (5) and (10) are favoured in the optics community. It is thus important to take great care when changing the constraint structure in the algorithms. As we have seen, while the form of prescriptions of projection algorithms in terms of fixed point iterations $u_{n+1} = \mathcal{V}(\mathcal{T}, \beta_n)u_n$ does not depend on the underlying constraints, this is not the case for prescriptions of the form (4), (5) and (10). This continues to cause some confusion in the optics literature where it is often assumed that simple changes in the constraint sets for the fixed-point prescriptions carries over in a straightforward manner to prescriptions of the form (4), (5) and (10). When written as fixed-point iterations, the effect of changing the constraint structure is seen in the mathematical properties of the operator rather than the form of the algorithm. To see how to translate the fixed-point algorithm (9) to the equivalent algorithm of the form (10), the reader should consult the proof of [4, proposition 2]. In our development of the RAAR algorithm below, we show the analogous analysis in order to translate the fixed-point algorithm into the more physically intuitive format.

Performance measures of the above algorithms are notoriously difficult to assess. The principal criteria we use are speed, quality of solutions and stability. Since each of the algorithms involves the same basic operations at each iteration, the speed of one algorithm relative to another boils down to iteration counts to convergence. But since, as we explain below, in practical settings these problems have no solution (that is, they are *infeasible*), the convergence is based on an error metric that is a rough indicator of image quality. To our knowledge there is no reliable objective metric for image quality. Ultimately, the quality comparison between algorithms is based on what looks good to the user. Nevertheless, the error metric we use is motivated by the convergence theory for fixed-point algorithms and is somewhat different from the conventional root-mean-squared (RMS) error metric used in the optics literature. Stability, however, can be objectively determined, and this clearly distinguishes the algorithms. By stability we mean that the algorithm reliably approaches a neighbourhood of a solution and remains there. A common and vexing problem for the HIO algorithm is that, if allowed to run too long, iterates will wander away from the neighbourhood of a solution, and the images will worsen. To circumvent this, practitioners have applied a variety of *ad hoc* recipes for applying the HIO algorithm, amounting to doing n steps of HIO, switching to another algorithm for m steps, switching back to HIO for another r steps, and stopping after running s steps of perhaps another algorithm. This problem of 'wandering' of the iterates, as we will explain below, is not unique to the HIO algorithm, and is rather a result of the nonconvexity of the constraints. For most optics applications, HIO, HPR and the difference map will all suffer from wandering, but preliminary numerical results indicate that the HPR algorithm is a promising alternative. Experiments on simulated data indicate that, once in the neighbourhood of a solution, the HPR iterates stay in that neighbourhood [4]. Moreover, by both the error metric and the subjective 'eye-ball norm', the images delivered by the HPR algorithm are superior to those of the HIO algorithm [4]. A drawback to the HPR algorithm is that, while it consistently delivers higher quality solutions than HIO, the path that the HPR algorithm takes to a solution can involve an intermediate stage in which the iterates are *further* from a solution, measured both by the error metric and the eye-ball norm, than iterates of the HIO algorithm at the same stage of the iteration. In other words, the HIO algorithm reaches a neighbourhood of a solution to the feasibility problem (1) in fewer iterations than HPR, but future HIO iterates will diverge from this neighbourhood; furthermore, the HPR iterates appear eventually to find a tighter neighbourhood of a solution than HIO iterates.

HPR has some modest analytical advantages as well. Detailed convergence results have been obtained in [5] for the unrelaxed HPR algorithm ($\beta = 1$) in a convex setting. In contrast to this, there are no convergence results available for the HIO algorithm applied to support and nonnegativity constraints—for any value of β—since no convex analogue to (4) has been found. At this time, however, there are no mathematically rigorous results proving convergence for the relaxed HPR algorithm ($\beta \neq 1$) or suggesting how to choose the relaxation parameter β to improve performance. Even less is known about the difference map which includes HPR as a special case.

We focus on the intermediate stage of these algorithms where HIO appears to outperform HPR. The algorithm we propose next achieves improved performance at the intermediate stage and superior stability/quality of the HPR algorithm at later iterations through an analytically motivated relaxation strategy.

2.3. Relaxed averaged alternating reflections

The new algorithm we propose is given by the following: given any $u_0 \in \mathcal{L}$, generate the sequence $u_0, u_1, u_2, \ldots$ by

$$u_{n+1} = V(\mathcal{T}_*, \beta_n)u_n \tag{11}$$

where

$$V(\mathcal{T}_*, \beta) = \beta\mathcal{T}_* + (1-\beta)P_M \qquad \text{and} \qquad \mathcal{T}_* = \tfrac{1}{2}(R_{S_+}R_M + I). \tag{12}$$

To underscore the connection of this algorithm with the averaged alternating reflection (AAR) algorithm studied in [5], we refer to (11) as the (RAAR) algorithm. For $\beta = 1$ the RAAR, HPR, AAR and the difference map ($\gamma_1 = -1$ and $\gamma_2 = 1/\beta$) algorithms are equivalent. For $\beta \neq 1$ the RAAR algorithm is fundamentally different from HPR; moreover, it cannot be derived as a special case of the difference map (8). The recursion (11) can be written analogously to (4) and (10). To see this, we proceed as in proposition 2 of [4]. Given an arbitrary signal $v \in \mathcal{L}$, let $v^+ = \max\{v, 0\}$ and $v^- = \min\{v, 0\}$ be its positive and negative parts, respectively. Then (11) can be rewritten as

$$u_{n+1} = (-\mathcal{X}_{D^c} \cdot \beta_n(2P_M - I) - [\mathcal{X}_D \cdot \beta_n(2P_M - I)]^- + P_M)(u_n). \tag{13}$$

There are three cases to consider: (i) If $x \in D$ and $(R_M u_n)(x) \geqslant 0$, then (13) yields $u_{n+1} = P_M$; (ii) if $x \in D$ and $(R_M u_n)(x) < 0$, then (13) becomes $u_{n+1}(x) = (((1-2\beta_n)P_M + \beta_n I)(u_n))(x)$; (iii) if $t \notin D$, then (13) can also be written as $u_{n+1}(x) = (((1-2\beta_n)P_M + \beta_n I)(u_n))(x)$. Altogether this yields the following algorithm:

$$(\forall x \in \mathbb{Z}^N) \qquad u_{n+1}(x) = \begin{cases} (P_M(u_n))(x), & \text{if} \quad x \in D \quad \text{and} \\ & (R_M(u_n))(x) \geqslant 0; \\ \beta_n u_n(x) - (1-2\beta_n)(P_M(u_n))(x), & \text{otherwise.} \end{cases} \tag{14}$$

We summarize the above discussion in the following proposition.

Proposition 2.1. *Algorithm (14) is equivalent to the recursion (11).*

The update rule in algorithm (14) depends on the pointwise sign of the reflector $(R_M(u_n))(x)$ whereas the update rule for Fienup's HIO algorithm (4) depends on the pointwise sign of the projector $(P_M(u_n))(x)$. The difference between the RAAR update rule and that for HPR (10) is much starker. Also note that the 'otherwise' action is simply a relaxation of the conditional action in the HIO algorithm; this is, again, very different from the HPR algorithm.

2.4. Convex analysis

To gain some insight into the behaviour of the algorithm above, we study the behaviour of the convex analogue to $V(\mathcal{T}_*, \beta)$. Let A and B be two closed convex subsets of $\mathcal{L}$. Replace S_+ and M by A and B respectively. Let $E \subset A$ denote the set of points in A nearest to B, and let $F \subset B$ denote the set of points in B nearest to A. The *gap vector* between A and B, denoted by $g \in \mathcal{L}$, is defined by $g = P_{\mathrm{cl}(B-A)}(0)$. Loosely interpreted, this is a vector pointing from E to F with $\|g\|$ measuring the smallest distance between A and B. For instance, if $A \cap B \neq \emptyset$

then $g = 0$. For a more precise treatment see [1, 2]. The convex counterpart to (12), the central operator in the RAAR algorithm, is defined by

$$V(T_*, \beta) = \beta T_* + (1-\beta)P_B, \qquad 0 < \beta < 1 \qquad \text{where} \qquad T_* = \tfrac{1}{2}(R_A R_B + I). \tag{15}$$

When discussing convergence of projection-type algorithms, one must take care to distinguish between *consistent* and *inconsistent* feasibility problems. In the current convex setting, consistent problems satisfy $A \cap B \neq \emptyset$; when $A \cap B = \emptyset$ the problem is said to be inconsistent. Inconsistent problems are common in applications where the *a priori* information represented by the constraint sets is highly idealized, particularly in the presence of noise. Bauschke *et al* [5] show that the properties of the AAR algorithm (that is, RAAR with $\beta = 1$) for consistent problems are very different from inconsistent problems. The reason for this is that the operator T_* *does not have* a fixed point if $A \cap B = \emptyset$. For $0 < \beta < 1$ the convex instance of the RAAR algorithm avoids these complications by transferring questions of consistency of the constraints to the existence of *nearest* points. In other words, the RAAR operator enjoys the advantage that $\mathrm{Fix}\, V(T_*, \beta)$ is independent of whether or not the associated feasibility problem is consistent. This is the content of the following theorem.

Theorem 2.2. *Let* $0 < \beta < 1$. *Then*

$$\mathrm{Fix}\, V(T_*, \beta) = F - \frac{\beta}{1-\beta} g \tag{16}$$

where g *is the gap vector between* A *and* B, *and* $F \subset B$ *is the set of points in* B *nearest to* A. *Moreover, for every* $u \in \mathrm{Fix}\, V(T_*, \beta)$, *we have the following:*

$$\begin{aligned} &(i) \quad u = P_B u - \frac{\beta}{1-\beta} g; \\ &(ii) \quad P_B u - P_A R_B u = g; \\ &(iii) \quad P_B u \in F \quad \text{and} \quad P_A P_B u \in E. \end{aligned} \tag{17}$$

Proof. To prove the result we must show (a) that $F - \beta g/(1-\beta) \subset \mathrm{Fix}\, V(T_*, \beta)$ and (b), conversely, that $\mathrm{Fix}\, V(T_*, \beta) \subset F - \beta g/(1-\beta)$. The first statement (a) is proved analogously to the proof of equation (18) of [5]. In the interest of brevity, we leave this as an exercise.

We show that $\mathrm{Fix}(\beta T_* + (1-\beta)P_B) \subset F - \frac{\beta}{1-\beta} g$. To see this, pick any $u \in \mathrm{Fix}(\beta T_* + (1-\beta)P_B)$. Let $f = P_B u$ and $y = u - f$. Recall that

$$P_A(2f - u) = P_A(2P_B u - u) = P_A R_B u. \tag{18}$$

This, together with the identity [5, proposition 3.3(i)]

$$(\forall u \in \mathcal{L}) \qquad u - T_* u = P_B u - P_A R_B u, \tag{19}$$

yields

$$P_A(2f - u) = f + T_* u - u. \tag{20}$$

For our choice of u we have $\beta T_* u + (1-\beta)P_B u = u$, which yields

$$T_* u - u = \frac{1-\beta}{\beta}(u - P_B u). \tag{21}$$

Then (20) and (21) give

$$P_A(2f - u) = f + \frac{1-\beta}{\beta}(u - f) = f + \frac{1-\beta}{\beta} y. \tag{22}$$

Now, for any $a \in A$, since A is nonempty, closed and convex, we have

$$\langle a - P_A(2f - u), (2f - u) - P_A(2f - u)\rangle \leqslant 0, \tag{23}$$

and hence

$$0 \geqslant \left\langle a - \left(f + \frac{1-\beta}{\beta} y \right), (2f - u) - \left(f + \frac{1-\beta}{\beta} y \right) \right\rangle$$
$$= \left\langle a - \left(f + \frac{1-\beta}{\beta} y \right), -y - \frac{1-\beta}{\beta} y \right\rangle$$
$$= \frac{1}{\beta} \langle -a + f, y \rangle + \frac{(1-\beta)}{(\beta)^2} \|y\|^2. \tag{24}$$

Here we have used (23), (22) and the fact that $y = u - f$. On the other hand, for any $b \in B$, since B is a nonempty closed convex set and $f = P_B u$, we have

$$\langle b - P_B u, u - f \rangle \leqslant 0, \tag{25}$$

which yields

$$\langle b - f, y \rangle = \langle b - f, u - f \rangle \leqslant 0. \tag{26}$$

Together, (24) and (26) yield

$$\langle b - a, y \rangle \leqslant -\frac{1-\beta}{\beta} \|y\|^2 \leqslant 0. \tag{27}$$

Now take a sequence $a_0, a_1, a_2, \ldots$ in A and a sequence $b_0, b_1, b_2, \ldots$ in B such that $g_n = b_n - a_n \to g$. Then

$$(\forall n \in \mathbb{N}) \qquad \langle g_n, y \rangle \leqslant -\frac{1-\beta}{\beta} \|y\|^2 \leqslant 0. \tag{28}$$

Taking the limit and using the Cauchy–Schwarz inequality yields

$$\|y\| \leqslant \frac{\beta}{1-\beta} \|g\|. \tag{29}$$

Conversely, $u - (\beta T_* u + (1-\beta) P_B u) = \beta(f - P_A(2f - u)) + (1-\beta) y = 0$ gives

$$\|y\| = \frac{\beta}{1-\beta} \|f - P_A(2f - u)\| \geqslant \frac{\beta}{1-\beta} \|g\|. \tag{30}$$

Hence $\|y\| = \frac{\beta}{1-\beta}\|g\|$ and, taking the limit in (28), $y = -\frac{\beta}{1-\beta} g$, which confirms (i). From (18) and (22) with $y = -\frac{\beta}{1-\beta} g$ it follows that $f - P_A R_B u = g$ which proves (ii) and, by definition, implies that $P_B u = f \in F$ and $P_A P_B u \in E$. This yields (iii) and proves (16). □

In words, regardless of whether or not $A \cap B$ is empty, as long as there are points in B that are nearest to A, then the RAAR operator $V(T_*, \beta)$ has a set of fixed points, and these are precisely the points in B nearest to A, translated by the scaled gap vector. This is the starting point for the convex heuristics behind the RAAR algorithm. Statements about convergence and more detailed behaviour of the algorithm are beyond the scope of this work.

We conclude the mathematical analysis with some observations that motivate the relaxation strategy we implement in section 3. We wish to use the parameter β to control the step size between successive iterates and, as much as possible, to steer the iterates. Far away from the solution, it is easy to see the damping effect of the parameter $0 < \beta < 1$, which derives from the form of the relaxation (12) as simply a convex combination of the operator $\mathcal{T}_*$ and the projector onto the *data* P_M—the smaller the relaxation parameter β, the closer to the data we require the iterates to stay. It was noted in [4] that, regardless of the relaxation, the HPR algorithm (10) takes significantly longer than the HIO algorithm (4) to reach a suitable neighbourhood of the solution, although, once near a solution, HPR delivers consistently better images with greater stability and reliability than HIO. We show in the following section that the dampening effect of the relaxation in the RAAR algorithm is just what is needed to control the initial behaviour of the HPR algorithm.

For the behaviour of the algorithm near the solution, we rely on the convex analysis. By (16), the relaxation parameter β affects the fixed points of the operator through the gap vector. If the feasibility problem is consistent, that is, $A \cap B \neq \emptyset$, then the gap vector $g = 0$. In this case, is it not clear what effect, if any, β will have on convergence. On the other hand, if the problem is inconsistent, that is, $A \cap B = \emptyset$, and $g \neq 0$, then, by (16), the set of nearest points F can be translated arbitrarily far away in the direction g by letting β approach 1 from below. We use this to gain some control on the step size between successive iterates and the directions of the steps.

Proposition 2.3. *Let $u_n \in \mathcal{L}$ satisfy $\|u_n - u_{\beta_n}\| < \delta$ where $u_{\beta_n} \in \mathrm{Fix}\, V(T_*, \beta_n)$ and $V(T_*, \beta_n)$ is the RAAR relaxation strategy defined by (15) with $0 < \beta_n < 1$. Define $u_{n+1} = V(T_*, \beta_{n+1})u_n$ for any $0 < \beta_{n+1} < 1$. Then*

$$\left\| u_{n+1} - \left(f_{\beta_n} - \frac{\beta_{n+1}}{1-\beta_n} g \right) \right\| < \delta, \qquad \text{where} \qquad f_{\beta_n} = P_B u_{\beta_n} \in F. \tag{31}$$

Proof. For any $u \in \mathcal{L}$, by (19) and (15), we have $V(T_*, \beta_{n+1})u - V(T_*, \beta_n)u = (\beta_{n+1} - \beta_n)(P_A - I)R_B u$, which, together with (17) (i), yields

$$u_{\beta_n} - V(T_*, \beta_{n+1})u_{\beta_n} = \frac{\beta_{n+1} - \beta_n}{1-\beta_n} g, \qquad \text{or} \qquad V(T_*, \beta_{n+1})u_{\beta_n} = f_{\beta_n} - \frac{\beta_{n+1}}{1-\beta_n} g. \tag{32}$$

Since $V(T_*, \beta_{n+1})$, being a convex combination of nonexpansive operators, is itself nonexpansive, the result follows from (32). □

While the HPR algorithm gives quite stable solutions eventually, the above theory suggests that this stability can be improved in a controlled fashion. Consider the fixed-point iteration as a descent algorithm minimizing some error metric (in fact, minimizing the gap distance) where $-g$ is the direction of descent. By (31) and the first equation in (32),

$$u_{n+1} \approx V(T_*, \beta_{n+1})u_{\beta_n} = u_{\beta_n} - \frac{\beta_{n+1} - \beta_n}{1-\beta_n} g,$$

thus one can use β_{n+1} to affect steps in the direction $-g$ ranging, in the limit, from length $-\beta_n/(1-\beta_n)$ to 1 as β_{n+1} varies from 0 to 1 respectively. The difference $u_{n+1} - u_n$ for the unrelaxed algorithm ($\beta = 1$) was shown in [5] to converge to the negative gap vector $-g$ in the inconsistent case. The effect of the relaxation is primarily to dampen the iteration in the neighbourhood of a solution in the case of inconsistent problems. To see the advantage of this, consider the nonconvex case and suppose that the problem is inconsistent (that is, the gap vector $g \neq 0$). The only case of the HPR algorithm for which we can say anything is the case $\beta = 1$, which is the same as the unrelaxed RAAR (or AAR) algorithm, so we restrict the discussion to the RAAR and AAR algorithms. The convex analysis of the AAR algorithm shows that, even though the gap is attained, the iterates u_n continue to move in the direction $-g$ without end. In the nonconvex setting, even if the true gap is attained, the continued progress of the iterates in the direction $-g$ could push the iterates away from the domain of attraction of the local solution and into a different domain of attraction. Thus the projections of the iterates, or the *shadows* might never converge. This 'wandering' of the iterates near an apparent local solution has been observed both with the HIO and HPR algorithms, though it is much less severe and destabilising with HPR than it is with HIO. The relaxations in the RAAR algorithm can be used to either dampen the iterates near a local solution to slow drifting out

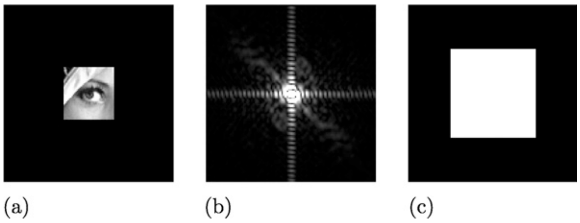

Figure 1. Original images and corresponding data used for the comparison of the HIO and HPR algorithms. The centre of (a) is a 38×38 pixel section of the standard Lena image, zero-padded to a 128×128 matrix. Frame (b) is the noiseless Fourier magnitude data m corresponding to image (a). The same object domain support constraint (and initial guess) of size 64×64 pixels, shown in (c), is used for each trial.

of a domain of attraction, or to halt the wandering of the iterates altogether by holding the relaxation parameter at a fixed value less than 1.

3. Numerical implementation

Our goal with the RAAR algorithm is to use dynamic relaxations to shorten the initial 'warm-up' phase of the HPR algorithm and to stabilize the algorithm near a local solution. The algorithm we consider is

$$u_{n+1} \approx V(\mathcal{T}_*, \beta_n)u_n. \tag{33}$$

Before outlining our specific implementation, some remarks are in order about the calculation of $\mathcal{T}$ given by (12). As discussed in [15, Section 5.2], the projection onto the magnitude constraint P_M is a numerically unstable operation due to the multivaluedness of the projection operator. We therefore recommend the following approximation to P_M (see [15, equation (74)]):

$$P_M u \approx \nabla J_\epsilon u = I - \mathcal{F}^{-1}\left(\left(\frac{|\mathcal{F}u|^2}{(|\mathcal{F}u|^2+\epsilon^2)^{1/2}} - m\right)\frac{|\mathcal{F}u|^2+2\epsilon^2}{(|\mathcal{F}u|^2+\epsilon^2)^{3/2}}\mathcal{F}u\right) \tag{34}$$

for $0 < \epsilon \ll 1$, where

$$J_\epsilon(u) = \frac{1}{2}(\|u\|^2 - \|\mathcal{F}^{-1}\widehat{v} - m\|^2), \qquad \text{where} \qquad \widehat{v} = \frac{|\mathcal{F}u|^2}{(|\mathcal{F}u|^2+\epsilon^2)^{1/2}}. \tag{35}$$

Define

$$\widetilde{\mathcal{T}}_* = \tfrac{1}{2}(R_{S_+}(2\nabla J_\epsilon - I) + I). \tag{36}$$

Under reasonable assumptions, by the continuity of R_{S_+} and [15, corollary 5.3], it can be shown that $\nabla J_\epsilon(u) \to P_M(u)$ and $V(\widetilde{\mathcal{T}}_*, \beta)u \to V(\mathcal{T}_*, \beta)u$ as $\epsilon \to 0$. For our experiments, we choose ϵ to be an order of magnitude larger than machine zero (double precision) relative to the square of the largest data element, that is, $\epsilon = 10^{-15}\max(|(\mathcal{F}u)|^2)$.

Using the stable approximation $V(\widetilde{\mathcal{T}}_*, \beta)$ given by (36), from the initial guess u_0 we generate the sequence $u_0, u_1, u_2, \ldots$ by

$$u_{n+1} = V(\widetilde{\mathcal{T}}_*, \beta_n)u_n \qquad \text{where} \qquad \beta_{n+1} = \beta_0 + (1-\beta_0)(1-\exp(-(n/7)^3)). \tag{37}$$

The rule for updating β_n is a smooth approximation to a step function from the value β_0 to the value 1 centred at iteration $n = 7$. We compare this algorithm to the HIO (4) and HPR (10) algorithms using the same stable projection approximation. We study algorithm performance with noisy data. The initial points u_0 are chosen to be the normalized characteristic function of the support constraint shown in figure 1(c).

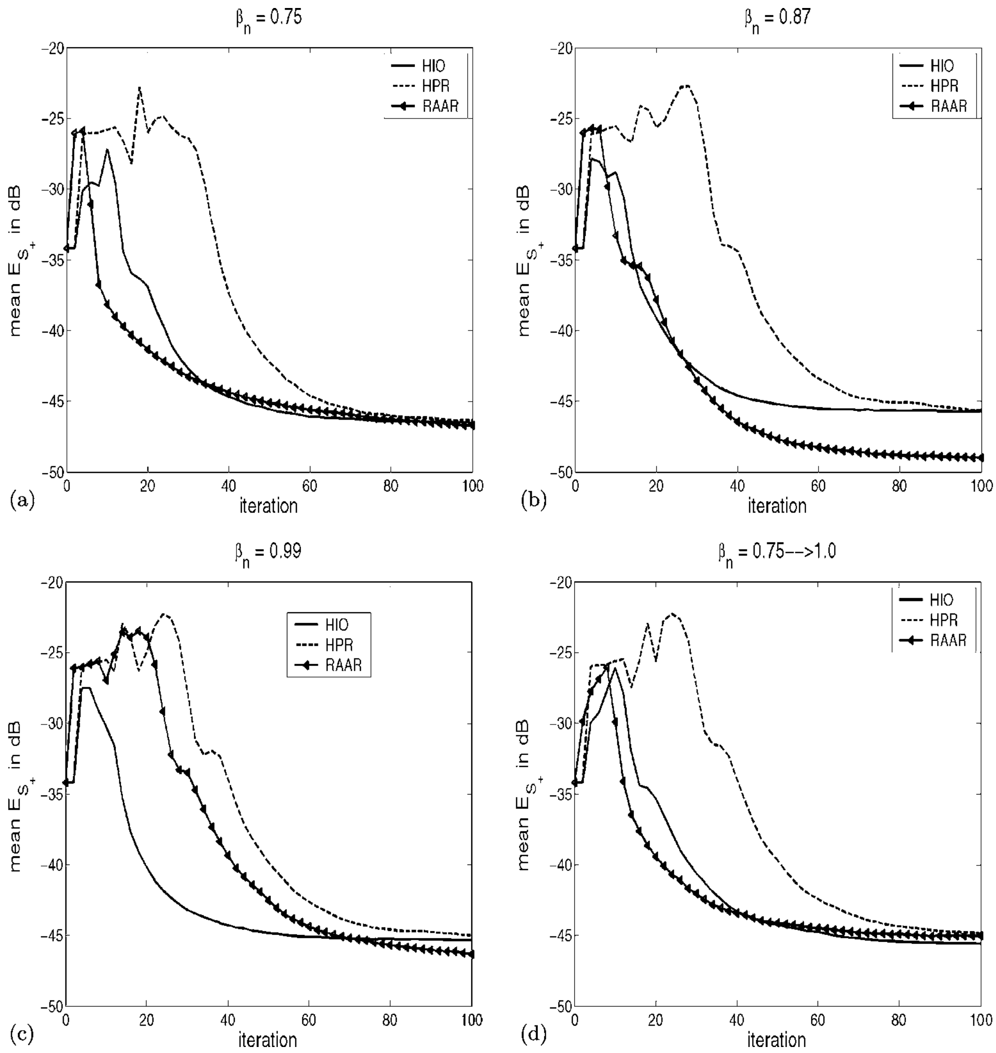

Figure 2. Error metric $E_{S_+}(x_n)$ averaged over 100 realizations of noise (SNR = 34 dB). For (a)–(c) the relaxation parameter for the respective algorithms, β_n, is fixed. For (d) β_n varies from 0.75 to 1.0 according to (37).

The data consist of the support/nonnegativity constraint, shown in figure 1(c), and Fourier magnitude data m, shown in figure 1(b), with additive noise η—a symmetric, randomly generated array with a zero mean Gaussian distribution. Following the experimental design of [4], the signal-to-noise ratio (SNR) is $20\log_{10}\|m\|/\|u\| = 34$ dB.

As in [4], the error metric we use to monitor the algorithms, E_{S_+}, is given by

$$E_{S_+}(x_n) = \frac{\left\| P_{S_+}(P_M(u_n)) - P_M(u_n)\right\|^2}{\|P_M(u_n)\|^2}. \tag{38}$$

Originally suggested by Fienup [10], this error metric is motivated from the observation that one is often more interested in signals satisfying the object domain constraint S_+. It is natural, then, to monitor the squared distance from the signal $P_M(u_n)$ to the set S_+, that is, $\left\| P_{S_+}(P_M(u_n)) - P_M(u_n)\right\|^2$. The denominator of (38) is just a normalization relative to the

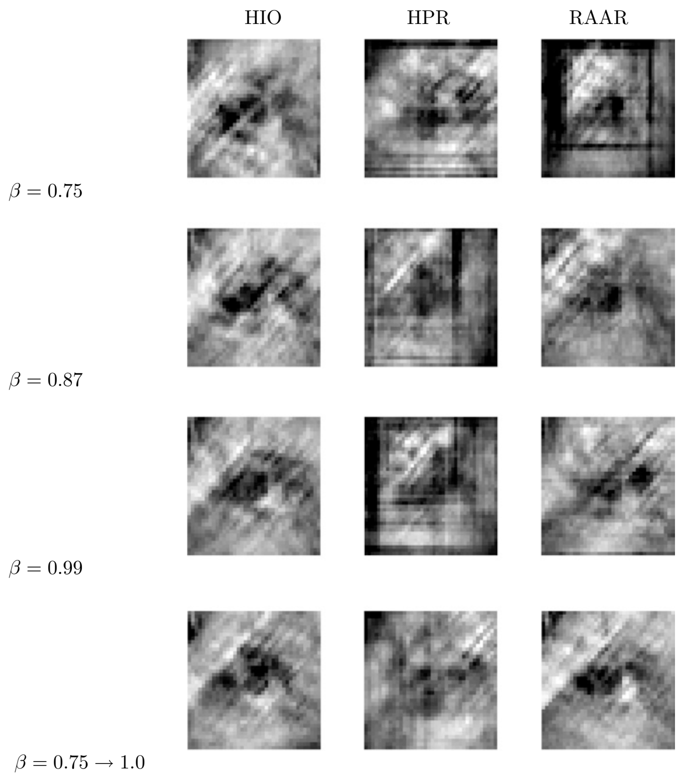

Figure 3. Typical images recovered after 35 iterations of the HIO, HPR and RAAR algorithms for different relaxation strategies with the same realization of data noise (SNR = 34 dB) and the same normalized initial guess. The variable β_n trials were generated according to the rule given by (37).

signal energy. It is important to note that this error metric is *in situ* in the sense that one need not know what the true object is. The root-mean-squared error favoured in the optics literature, in contrast, is given by

$$E_{\mathrm{RMS}}(u_n) = \frac{\||\mathcal{F}u_n| - m\|}{\|m\|}. \tag{39}$$

This monitors the distance of the iterate u_n to the data rather than the *a priori* information. A deeper mathematical motivation for (38) based on fixed-point theory can be found in [5].

We compute the mean value of the error measure E_{S+} over 100 trials with different realizations of the noise and the same initial guess.

First, we compare the mean behaviour over 100 iterations of four sets of realizations of the algorithms, each corresponding to different relaxation strategies, $\beta = 0.75$, $\beta = 0.87$, $\beta = 0.99$ and variable β_n governed by (37) with $\beta_0 = 0.75$. The average value of the error metric at iteration n, $E_{S_+}(x_n)$, is shown in figure 2. These are all given in decibels (recall that the decibel value of $\alpha > 0$ is $10\log_{10}(\alpha)$). In figure 3 we show typical estimates generated by the respective algorithms at iteration 35, all from the same realization of noise and the same initial guess. While the RAAR algorithm with $\beta = 0.75$ appears to perform well as measured by E_{S_+} (see figure 2(a)), it is clear from figure 3 that the quality of solutions found by the RAAR algorithm degrades rapidly as the relaxation parameter β becomes small. For values of β near 1.0 the quality of the iterates generated by the RAAR algorithm does eventually improve, however, as with the HPR algorithm, it takes many more iterations to achieve this improvement. For static values of β the best performance for the RAAR algorithm appears to be achieved with a value of $\beta = 0.87$. The variable β_n trials for the RAAR algorithm yielded the best overall results, measured both by the error metric, as well as observed picture quality. In contrast to this, the relaxation parameter does not appear to have any identifiable effect on the performance of the HIO or HPR algorithms.

4. Concluding remarks

There are infinitely many relaxation strategies one could implement for iterative transform methods, but very few of them admit a meaningful mathematical analysis. The standard for phase retrieval algorithms, Fienup's HIO algorithm, has been identified in a special case with the promising HPR algorithm, which in turn, has been identified as a special case of Elser's difference map. For each of these algorithmic frameworks, the mathematical properties of the algorithms vary dramatically with the parameter values in a manner analogous to bifurcations of dynamical systems. A complete mathematical analysis must treat all relevant intervals of parameter values on a case-by-case basis. No such analysis is available for the HIO, HPR or difference map algorithms. To circumvent these difficulties and to improve upon the HPR algorithm, we propose a simple relaxation, the RAAR algorithm, of a well-understood AAR algorithm. The relaxation is a convex combination of the AAR fixed point operator, and the projection onto the *data*. This intuitive framework is mathematically tractable and suggests an easy strategy for the choice of relaxation parameter that, moreover, improves algorithm performance. In contrast, it appears that similar relaxation strategies have little effect on either the HIO or the HPR algorithm. We cannot recommend a rule by which to select a static value of β—this depends on the data. Nevertheless, based on the results for the variable β_n trials, we propose the fairly generic dynamic relaxation strategy of (37) for getting the best performance from the RAAR algorithm. Here the algorithm is significantly relaxed in the early iterations, helping the algorithm to find a neighbourhood of the solution quickly while maintaining fidelity to the data, and then decreasing the relaxation (i.e. increasing β_n) in the neighbourhood of the solution to avoid stagnation at a poor local minimum. To stabilize iterates in the domain of attraction of a solution, a final fixed value of β close to, but less than, 1, say $\beta = 0.999\,99$ should be chosen. In a technical point, we also recommend a smooth perturbation of the magnitude projector (34) to improve the numerical stability of computing the projection onto magnitude constraints.

Acknowledgments

This work was supported by a post-doctoral fellowship from the Pacific Institute for the Mathematical Sciences at Simon Fraser University. The author would like to thank Veit Elser for pointing out the connection between the HPR algorithm and the difference map. Thanks also to the anonymous referees for their helpful comments. Special thanks to Heinz Bauschke and Patrick Combettes for their careful reading and valuable comments during the preparation of this work.

References

[1] Bauschke H H and Borwein J M 1993 On the convergence of von Neumann's alternating projection algorithm for two sets *Set-Valued Anal.* **1** 185–212

[2] Bauschke H H and Borwein J M 1994 Dykstra's alternating projection algorithm for two sets *J. Approx. Theory* **79** 418–43

[3] Bauschke H H, Combettes P L and Luke D R 2002 Phase retrieval, error reduction algorithm, and Fienup variants: a view from convex optimization *J. Opt. Soc. Am.* A **19** 1334–45
[4] Bauschke H H, Combettes P L and Luke D R 2003 Hybrid projection–reflection method for phase retrieval *J. Opt. Soc. Am.* A **20** 1025–34
[5] Bauschke H H, Combettes P L and Luke D R 2004 Finding best approximation pairs relative to two closed convex sets in Hilbert spaces *J. Approx. Theory* **127** 178–92
[6] Burke J V and Luke D R 2003 Variational analysis applied to the problem of optical phase retrieval *SIAM J. Control Optim.* **42** 576–95
[7] Elser V 2003 Phase retrieval by iterated projections *J. Opt. Soc. Am.* A **20** 40–55
[8] Elser V 2003 Solution of the crystallographic phase problem by iterated projections *Acta Crystallogr.* A **59** 201–209
[9] Faulkner H M L, Allen L J, Oxley M P and Paganin D 2003 Computational aberration determination and correction *Opt. Commum.* **216** 89–98
[10] Fienup J R 1982 Phase retrieval algorithms: a comparison *Appl. Opt.* **21** 2758–69
[11] Gerchberg R W and Saxton W O 1972 A practical algorithm for the determination of phase from image and diffraction plane pictures *Optik* **35** 237–46
[12] Hauptman H A 1991 The phase problem of x-ray crystallography *Rep. Prog. Phys.* **54** 1427–54
[13] He H, Marchesini S, Howells M, Weierstall U, Hembree G and Spence J 2003 Experimental lensless soft-x-ray imaging using iterative algorithms: phasing diffuse scattering *Acta Crystallogr.* A **59** 143–52
[14] Howells M R, Chapman H, Hau-Riege S, He H, Marchesini S, Spence J and W U 2003 X-ray microscopy by phase-retrieval methods at the advanced light source *J. Physique* IV **104** 557–61
[15] Luke D R, Burke J V and Lyon R J 2002 Optical wavefront reconstruction: theory and numerical methods *SIAM Rev.* **44** 169–224
[16] Millane R P 1990 Phase retrieval in crystallography and optics *J. Opt. Soc. Am.* A **7** 394–411
[17] Millane R P 1993 Phase problems for periodic images—effects of support and symmetry *J. Opt. Soc. Am.* A **10** 1037–45
[18] Millane R P and Stroud W J 1997 Reconstructing symmetric images from their undersampled Fourier intensities *J. Opt. Soc. Am.* A **14** 568–79
[19] Cloetens P, PateyronSalome M, Buffiere J Y, Peix G, Baruchel J, Peyrin F and Schlenker M 1997 Observation of microstructure and damage in materials by phase sensitive radiography and tomography *J. Appl. Phys.* **81** 5878–86
[20] Pagot E, Cloetens P, Fiedler S, Bravin A, Coan P, Baruchel J, Hartwig J and Thomlinson W 2003 A method to extract quantitative information in analyzer-based x-ray phase contrast imaging *Appl. Phys. Let.* **82** 3421–23
[21] Spence J C H, Wu J S, Giacovazzo C, Carrozzini B, Cascarano G L and Padmore H A 2003 Solving non-periodic structures using direct methods: phasing diffuse scattering *Acta Crystallogr.* A **59** 255–61
[22] Stevenson A W *et al* 2003 Phase-contrast x-ray imaging with synchrotron radiation for materials science applications *Nucl. Instrum. Methods* B **199** 427–35