LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

UCRL-JRNL-219850

A unified evaluation of iterative projection algorithms for phase retrieval

S. Marchesini

March 16, 2006

Review of Scientific Instruments

Disclaimer

A unified evaluation of iterative projection algorithms for phase retrieval

S. Marchesini[1]

[1]*Lawrence Livermore National Laboratory, 7000 East Ave., Livermore, CA 94550-9234, USA*

(Dated: March 8, 2006)

Iterative projection algorithms are successfully being used as a substitute of lenses to recombine, numerically rather than optically, light scattered by illuminated objects. Images obtained computationally allow aberration-free diffraction-limited imaging and allow new types of imaging using radiation for which no lenses exist. The challenge of this imaging technique is transfered from the lenses to the algorithms. We evaluate these new computational "instruments" developed for the phase retrieval problem, and discuss acceleration strategies.

Crystallographers routinely image molecular structures of several thousand atoms by phasing the diffraction pattern of a structure replicated in a periodic system. Likewise, computationally retrieving the phase of a diffraction pattern is becoming increasingly successful at imaging -with several millions of resolution elements- isolated objects with no translational symmetry as complex as as biological cells, nanotubes and nanoscale aerogel structures. Diffraction microscopy (the imaging of isolated objects by diffraction and computational phase retrieval) promises a 3D resolution limited only by radiation damage, wavelength, the collected solid angle and the number of x-rays or electrons collected. This capability provides an extremely valuable tool for understanding nanoscience and cellular biology. Recent estimates [1] of the dose and flux requirements of x-ray diffraction on single objects, indicate that attractive resolution values (about 10 nm for life science and 2–4 nm for material science) should be possible at a modern synchrotron. Atomic resolution could be accomplished using pulses of x-rays that are shorter than the damage process itself [2] using femtosecond pulses from an x-ray free-electron laser [3]. Alternatively the radiation damage limit could be eliminated by continuously replacing the exposed samples with identical ones such as laser aligned molecules [4].

In the fields of electron microscopy [5] and astronomical imaging [6], iterative projection algorithms have been used to recover the phase information in a variety of problems. The evaluation of the aberrations in the Hubble space telescope by Fienup in [7] remains perhaps the most prominent example of successful phase reconstructions in the astronomical community. In electron diffraction microscopy [5, 8] Zuo and coworkers imaged a single isolated nanotube at atomic resolution [9], Wu et al. imaged defects at atomic resolution [10].

An important review, which attempted to integrate the approaches of the optical and crystallographic communities, appeared in 1990 [11]. The connection was made between the "solvent-flattening" or "density-modification" techniques of crystallography [12] and the compact support requirements of the iterative projection algorithms. The importance of fine sampling of the *intensity* of the measured diffraction pattern was recognized at an early stage [13].

The observation by Sayre in 1952 [14] that Bragg diffraction under-samples the diffracted intensity pattern was important and led to more specific proposals by the same author for X-ray diffractive microscopy of non-periodic objects [15, 16]. These ideas, combined with the rapid development of computational phase retrieval in the wider optics community especially the "support constraint" [5, 6, 18, 19], enabled the first successful use of Coherent X-ray Diffraction Microscopy.

Since the first proof of principle demonstration of Coherent X-ray Diffraction Microscopy (CXDM) by a team in Stony Brook [20], a number of groups have been working to bring these possibilities into reality.

Robinson and coworkers at the University of Illinois have applied the principles of CXDM to hard x-ray experiments on micro-crystalline particles. Such data have been reconstructed tomographically to produce a 3D image at 80 nm resolution [21]. Miao (now at UCLA) and coworkers have made considerable progress in pushing the CXDM method in Spring-8 Japan to higher resolution in 2D (7 nm), higher x-ray energies and to a limited form of 3D [22]. They have also made the first application of CXDM to a biological sample [23].

A diffraction chamber dedicated to diffraction microscopy [24] has been used to image biological cells [25, 26] at the Advanced Light Source in Berkeley [27]. Using the same chamber a collaboration between Berkeley, Livermore labs and Arizona State University produced 3D imaging at 10x10x40 nm resolution of test samples [28] as well as aerogel foams [29].

In this article the computational instruments that enabled these and other results are reviewed. Section I introduces the phase problem and the experimental requirements for diffraction microscopy, Section II describes the concepts of sets of images and their projectors. In section III the iterative projection algorithms published in the literature are summarized and tested on simple geometric sets. In section IV the connection between projection and gradient based methods and related acceleration strategies are discussed.

I. THE PHASE PROBLEM

When we record the diffraction pattern scattered by an object the phase information is missing. Apart from normalization factors, an object with density $\rho(\boldsymbol{r})$, $\boldsymbol{r}$ being the coordinates in the *object* (or *real*) space, generates a diffraction pattern equal to the modulus square of the Fourier Transform (FT) $\tilde{\rho}(\boldsymbol{k})$:

$$\begin{aligned} I(\boldsymbol{k}) &= |\tilde{\rho}(\boldsymbol{k})|^2 \\ I(\boldsymbol{k}) &= \tilde{\rho}^{\dagger}(\boldsymbol{k})\tilde{\rho}(\boldsymbol{k})\,. \end{aligned} \tag{1}$$

Where $\boldsymbol{k}$ represent the coordinate in the Fourier (or Reciprocal) space. The inverse Fourier transform (IFT) of the measured intensity I provides the autocorrelation $\rho(-\boldsymbol{r}) * \rho(\boldsymbol{r})$ of the object:

$$\mathrm{IFT}[I(\boldsymbol{k})] = \rho(-\boldsymbol{r}) * \rho(\boldsymbol{r}) \tag{2}$$

The phase retrieval problem consists of solving $\tilde{\rho}$ in Eq. 1 or ρ in Eq. 2 using some extra prior knowledge.

Since the intensity represent the FT of the autocorrelation function, and the autocorrelation is twice as big as the object, the intensity should be sampled at least twice as finely as the amplitude to capture all information of the object. Finer sampling adds a 0-padding region around the recovered autocorrelation function which adds no further information (Shannon theorem). Less then critical sampling in the Fourier domain causes aliasing in the object space. A periodic repetition of the same structure provides stronger signal, enabling the measurement of the diffraction pattern before the structure is damaged. However, while an isolated object generates a continuous diffraction pattern that can be sampled as finely as desired, a periodic repetition of the same object generates only a subset of the possible diffraction intensities. Crystallography therefore has to deal with an aliased autocorrelation function, also known as the Patterson function. This reduced information can be compensated by other prior knowledge such as the atomistic nature of the object being imaged, knowledge of a portion of the object, presence of heavy atoms and information obtained with anomalous diffraction. Other information includes the presence of a solvent in the crystal. By varying the sampling rate of a diffraction pattern it was shown [16, 17] that less then critical sampling was sufficient to solve the phase problem. This was possible because the number of equations (measured intensities in eq. 1) in the two and three dimensional phase retrieval problem is larger than the number of unknowns (resolution elements in the object). The number of unknowns defines the number of independent equations, or minimum required sampling rate. Although no general proof has been provided that limited sampling removes only redundant equations, such a minimum required sampling rate suggest that when the solvent exceeds 50% of the crystal volume, the algorithms developed in the optical community using techniques to dynamically refine the solvent regions [30] may be able obtain ab-initio structural information from crystals.

Coherence is required to properly sample the FT of the autocorrelation of the object [31]. According to Schell theorem [32], the autocorrelation of the illuminated object obtained from the recorded intensity is multiplied by the complex degree of coherence. The beam needs to fully illuminate the isolated object, and the degree of coherence must be larger than the autocorrelation of the illuminated object.

Diffraction microscopy solves the phase problem using the knowledge that the object being imaged is isolated, it is assumed to be 0 outside a region called support S:

$$\rho(\boldsymbol{r}) = 0, \text{ if } \boldsymbol{r} \notin S \tag{3}$$

This support is equivalent to the solvent in crystallography. Equations 1 and 3 can be combined to obtain a multidimensional system of quadratic equations in the $\rho(\boldsymbol{r})$ variables:

$$\left| \sum_{\boldsymbol{r} \in S} \rho(\boldsymbol{r}) \exp(i\boldsymbol{k} \cdot \boldsymbol{r}) \right|^2 = I(\boldsymbol{k}) \tag{4}$$

which can be written as:

$$\sum_{\boldsymbol{r},\boldsymbol{r'} \in S} \rho(\boldsymbol{r})\rho^*(\boldsymbol{r'}) \cos(\boldsymbol{k} \cdot (\boldsymbol{r} - \boldsymbol{r'})) \;=\; I(\boldsymbol{k}) \tag{5}$$

Each value $I(\boldsymbol{k})$ in reciprocal space defines an ellipsoid (Eq. 5) in the multidimensional space of the unknowns $\rho(\boldsymbol{r})$, $\{\boldsymbol{r} \in S\}$. If the number of *independent* equations equals the number of unknowns the system has a single solution $\rho(\boldsymbol{r})$. The intersection of these ellipsoids forms our solution. Unfortunately this system of equations is difficult to solve, and has an enormous amount of local minima. Constant phase factors, inversion with respect to the origin (enantiomorphs) and origin shifts $\rho(\pm\boldsymbol{r} + \boldsymbol{r_0})e^{i\phi_0}$ are undetermined and considered equivalent solutions. The presence of multiple non-equivalent solutions in two and higher dimensional phase retrieval problems is rare [33] and occours when the density distribution of the object can be described as the convolution of two or more non-centrosymmetric distributions. Simple homometric structures for which the phase problem is not unique [34] exist in nature, but such non-uniqueness is less likely for more complex structures.

Presence of noise and limited prior knowledge increases the number of solutions within the noise level. Confidence that the recovered image is the correct and unique one can be obtained by repeating the phase retrieval process using several random starts. Repeatibility of the recovered images as a function of resolution measures the effective phase retrieval transfer function [25, 28], which can be decomposed in unconstrained amplitudes modes [25] and phase aberrations [35].

The development of iterative algorithms with feedback in the early nineteen-eighties by Fienup produced a remarkably successful optimization method capable of extracting phase information [6, 19, 36]. The important

theoretical insight that these iterations may be viewed as projections in Hilbert space [37, 38] has allowed theoreticians to analyze and improve on the basic Fienup algorithm [39, 40, 41, 42].

These algorithms try to find the intersection between two sets, typically the set of all the possible objects with a given diffraction pattern (modulus set), and the set of all the objects that are constrained within a given area or volume called support (or outside a solvent region in crystallography). The search for the intersection is based on the information obtained by *projecting* the current estimate on the two sets. An error metric is obtained by evaluating the distance between the current estimate and a given set. The error metric and its gradient are used in conjugate gradient (CG) based methods such as SPEDEN [43].

II. SETS, PROJECTORS AND METRICS

An image of a density distribution can be described as a sequence of n pixel values. For an image of n pixels, there are n coordinates. The magnitude of the density at a pixel defines the value of that coordinate. Thus a single vector in this n-dimensional space defines an image. For comples images the number of coordinates increases by a factor of two. Axes of the multidimensional space are formed by any sequence of n-pixels with all but one pixel equal to 0. For example $\boldsymbol{x} = (x, 0, 0)$ in 3-pixel solution space. The origin of this space is the image with all the pixels equal to 0. The components on these axes form the *real* or *object* space. The same object can be described in terms of any another n-dimensional orthogonal (or linearly independent) bases. Axes can be rotated, shifted, inverted and so on and the proper linear transform must be applied to obtain the components in the new basis. The basis used to describe the image must have at least n components, but more can be used if it helps to describe the properties of the algorithm. For example one could let the pixel values have a real and imaginary component, doubling the number of dimensions used to describe the object.

One important basis is the *momentum* or *Fourier* space. While the vector in the n-dimensional space representing an image is unaltered on transforming from real to reciprocal space, its components in the new axes are altered (Fourier transformed). The distance between two points in the n-dimensional space is independent of this trasformation (Parseval theorem). The lengths and the angles between vectors will be our guide to describe the behavior, convergence and error metrics of these algorithms.

We consider two sets, S (support) and M (modulus). When the image belongs to both sets simultaneously, we reached an equivalent solution. If the properties of the object being imaged are known a-priori to be limited in a certain region called support, we know that of all the possible values in the n-dimensional space, the pixel values, some of them must be zero. Images that satisfy this rule (Eq. 3) form the support constraint set. A projection onto this set ($\boldsymbol{P}_S$) involves setting to 0 the components outside the support, while leaving the rest of the values unchanged (Fig. 1(a)):

$$\boldsymbol{P}_s\rho(\boldsymbol{r}) = \begin{cases} \rho(\boldsymbol{r}) & \text{if } \boldsymbol{r} \in S \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

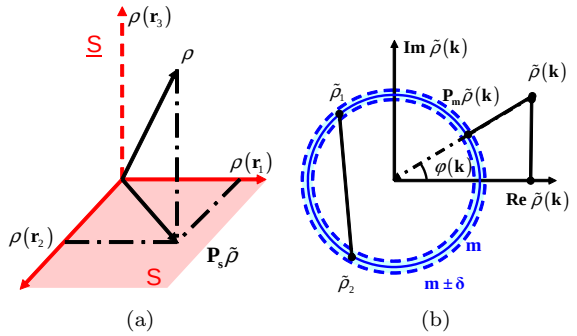


FIG. 1: Examples of sets and projectors: (a) Support: The axes represent the values on 3 pixels of an image ρ known to be 0 outside the support. The vertical axis $\rho(\boldsymbol{r}_3)$ represents a pixel outside the support ($\boldsymbol{r}_3 \in \underline{S}$), while the horizontal plane represents pixels inside the support S. The projection on this set is performed simply by setting to 0 all the pixels outside the support. (b) Modulus: A pixel (in Fourier space) with a given complex value is projected on the closest point on the circle defined by the radius m. If there is some uncertainty in the value of the radius $m \pm \delta$, the circle becomes a band. The circle is a non-convex set, since the linear combination between two points on the same set, ρ_1 and ρ_2 does not lie on the set. Also represented in the figure is the projection on the real axis (Reality projection).

The values in every pixel in Fourier space can be described using two components, the real and imaginary parts, or amplitude and phase, both defining a point in a complex plane. In an intensity measurement we obtain the amplitude or *modulus* in every pixel which defines a circle in a complex plane. These circles define the modulus constraint (1(b)). When every complex valued pixel lies on the circle defined by the corresponding modulus, the image satisfies this constraint and it belongs to the set. Chords - segments joining two points on a circle - do not belong to the circle, therefore linear combination of two images with all the pixel values lying on the same complex circles does not satisfy the constraint: the set is non-convex. Non-convex sets are problematic due to the presence of local minima and undefined projections.

The projection of a point in each complex plane onto the corresponding circle is accomplished by taking the point on the circle closest to the current one, setting the modulus to the measured one $m(\boldsymbol{k}) = \sqrt{I(\boldsymbol{k})}$, and leaving the phase unchanged (Fig. 1(b)):

$$\boldsymbol{P}_m\tilde{\rho}(\boldsymbol{k}) = \boldsymbol{P}_m|\tilde{\rho}(\boldsymbol{k})|e^{i\varphi(\boldsymbol{k})} = m(\boldsymbol{k})e^{i\varphi(\boldsymbol{k})}\,. \tag{7}$$

This operator is demonstrated to be a projector on the *non-convex* (Fig. 1(b)) set of the magnitude constraint [44]. The same paper discusses the problems of multivalued projections for non-convex sets, which do not satisfy the requirements for gradient-based minimization algorithms, and the related nonsmoothness of the squared set distance metric, which may lead to numerical instabilities. See also [45] for a follow-up discussion on the non-smooth analysis.

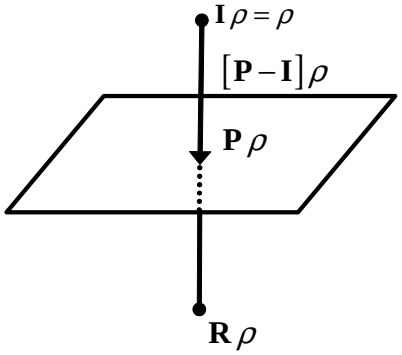


FIG. 2: The reflector applies the same step as the projector ($\boldsymbol{P}-\boldsymbol{I}$) twice: $\boldsymbol{R}\rho = \boldsymbol{I}\rho + 2[\boldsymbol{P}-\boldsymbol{I}]\rho$

A projector $\boldsymbol{P}$ is an operator that takes to the closest point of a set from the current point ρ. A repetition of the same projection is equal to one projection alone ($\boldsymbol{P^2} = \boldsymbol{P}$), therefore its eigenvalues must be $\lambda = 0, 1$. Another operator used here is the reflector $\boldsymbol{R} = \boldsymbol{I} + 2[\boldsymbol{P}-\boldsymbol{I}] = 2\boldsymbol{P} - \boldsymbol{I}$, which applies the same step as the projector but moves twice as far. In the case of the support constraint, the whole image space can be described in terms of the eigenvectors of the corresponding *linear* projector. These eigenvectors with eigenvalues of 1 (0) are the images with all the pixel equal to 0 except for one pixel inside (outside) the support. The modulus projector however is a nonlinear operator:

$$\begin{aligned} P_m(a+b) &\neq P_m(a) + P_m(b) \\ P_m(\alpha a) &\neq \alpha P_m(a) \end{aligned} \tag{8}$$

and it cannot be described in terms of eigenvalues and eigenvectors.

The Euclidean length $||\rho||$ of a vector ρ is defined as:

$$||\rho||^2 = \rho^\dagger \cdot \rho = \sum_{\boldsymbol{r}} |\rho(r)|^2 = \sum_{\boldsymbol{k}} |\tilde{\rho}(k)|^2. \tag{9}$$

The sum is extended to the measured portion of the diffraction pattern. If part of the reciprocal space is not measured, it should not be included in the sum. In fact the sum should be waited with the experimental noise $\sigma(\boldsymbol{k})$:

$$||\rho||^2 = \frac{\sum_{\boldsymbol{k}} \frac{1}{\sigma^2(\boldsymbol{k})} |\tilde{\rho}(k)|^2}{\sum_{\boldsymbol{k}} \frac{1}{\sigma^2(\boldsymbol{k})}}. \tag{10}$$

with $\sigma(\boldsymbol{k}) = \infty$ for values of $\boldsymbol{k}$ not measured. The distance from the current point to the set $||\boldsymbol{P}\rho - \rho||$ is the basis for our error metric. Typically the error in real (ε_s) and reciprocal space (ε_m) are defined in terms of their distance to the corresponding set:

$$\begin{aligned} \varepsilon_s(\rho) &= ||\boldsymbol{P}_s\rho - \rho||\,, \\ \varepsilon_m(\rho) &= ||\boldsymbol{P}_m\rho - \rho||\,, \end{aligned} \tag{11}$$

or their normalized version $\bar{\varepsilon}_x(\rho) = \frac{\varepsilon_x(\rho)}{||\boldsymbol{P}_x\rho||}$. Another error metric used in the literature is given by the distance between the two sets: $\varepsilon_{s,m}(\rho) = ||\boldsymbol{P}_m\rho - \boldsymbol{P}_s\rho||$. The projector $\boldsymbol{P}_m$ moves ρ to the closest minimum of $\varepsilon_m^2(\boldsymbol{P}_m\rho) = 0$, providing a simple relation with the gradient $\nabla_\rho \varepsilon_m^2(\rho)$ [19, 44]:

$$\boldsymbol{P_m}\rho = \rho + [\boldsymbol{P_m} - \boldsymbol{I}]\rho = \rho - \tfrac{1}{2}\nabla_\rho \varepsilon_m^2(\rho)\,, \tag{12}$$

where $\nabla_\rho \varepsilon_m^2(\rho)$ is proportional to $\nabla_\rho \varepsilon_m(\rho)$:

$$\nabla_\rho \varepsilon_m^2(\rho) = 2\varepsilon_m(\rho)\nabla_\rho \varepsilon_m(\rho). \tag{13}$$

For $\tilde{\rho}(\boldsymbol{k}) = 0$, ε_m is non differentiable, and the projector $\boldsymbol{P}_m$ is multivalued [44]. The presence of complex zeros ($\rho(\boldsymbol{k}) = 0$) is considered of fundamental importance to the phase retrieval problem [46], and the phase vortices associated with these zeros cause stagnation in iterative algorithms [47].

Similarly the projector P_s minimizes the error ε_s^2:

$$[\boldsymbol{I} - \boldsymbol{P}_s]\rho = \tfrac{1}{2}\nabla_\rho \varepsilon_s^2(\rho) \tag{14}$$

III. ITERATIVE PROJECTION ALGORITHMS

Several algorithms based on these concepts have now been proposed and a visual representation of their behavior is useful to characterize the algorithm in various situations, in order to help chose the most appropriate one for a particular problem. In this section the projection algorithms published in the literature are summarized (see also Tab. I) and tested on simple geometrical sets.

The following algorithms require a starting point ρ^0, which is generated by assigning a random phase to the measured object amplitude (modulus) in the Fourier domain $|\tilde{\rho}(\boldsymbol{k})| = m(\boldsymbol{k}) = \sqrt{I(\boldsymbol{k})}$. The first algorithm called *Error Reduction* (ER) (Gerchberg and Saxton [5]) (see also Alternating Projections Onto Convex Sets [48] or Alternating Projections Onto Nonconvex Sets [37]) is simply (Fig. 3(a)):

$$\rho^{(n+1)} = \boldsymbol{P_s P_m}\rho^{(n)}\,, \tag{15}$$

by projecting back and forth between two sets, it converges to the local minimum. The name of the algorithm is due to the steps moving along the gradient of the error metric (see Eq. 12):

$$\boldsymbol{P}_s\boldsymbol{P}_m\rho = \boldsymbol{P}_s\rho - \tfrac{1}{2}\nabla_{\rho\in S}\,\varepsilon_m^2(\rho)\,, \tag{16}$$

TABLE I: Summary of various algorithms

Algorithm	Iteration $\rho^{(n+1)} =$
ER or Solvent Flattening	$\boldsymbol{P_s P_m} \rho^{(n)}$
Solvent Flip	$\boldsymbol{R_s P_m} \rho^{(n)}$
HIO	$\begin{cases} \boldsymbol{P_m} \rho^{(n)}(\boldsymbol{r}) & \boldsymbol{r} \in S \\ (\boldsymbol{I} - \beta \boldsymbol{P_m}) \rho^{(n)}(\boldsymbol{r}) & \boldsymbol{r} \notin S \end{cases}$
Difference Map	$\{\boldsymbol{I} + \beta \boldsymbol{P_s} \; [(1+\gamma_s)\boldsymbol{P_m} - \gamma_s \boldsymbol{I}] - \beta \boldsymbol{P_m} \; [(1+\gamma_m)\boldsymbol{P_s} - \gamma_m \boldsymbol{I}]\} \rho^{(n)}$
ASR	$\frac{1}{2}[\boldsymbol{R_s R_m} + \boldsymbol{I}] \rho^{(n)}$
HPR	$\frac{1}{2}[\boldsymbol{R_s}\,(\boldsymbol{R_m} + (\beta-1)\boldsymbol{P_m}) + \boldsymbol{I} + (1-\beta)\boldsymbol{P_m}] \rho^{(n)}$
RAAR	$\left[\frac{1}{2}\beta\,(\boldsymbol{R_s R_m} + \boldsymbol{I}) + (1-\beta)\boldsymbol{P_m}\right] \rho^{(n)}$

where $\nabla_{\rho \in S}\, \varepsilon_m^2(\rho)$ is the component of the gradient in the support. Fig. 3(a) shows that the step size is far from optimum, but guarantees linear convergence. A line search along this gradient direction would considerably speed up the convergence to a local minimum and will be discussed in IV

The *solvent flipping* algorithm [49] is obtained by replacing the support projector $\boldsymbol{P_s}$ with its reflector $\boldsymbol{R_s} = 2\boldsymbol{P_s} - \boldsymbol{I}$, one obtains (Fig. 3(b)) :

$$\rho^{(n+1)} = \boldsymbol{R_s P_m} \rho^{(n)} \,. \tag{17}$$

which multiplies the charge density ρ outside the support by -1. The *Hybrid Input Output* (HIO) [6, 19] (Fig. 3(c) is:

$$\rho^{(n+1)}(x) = \begin{cases} \boldsymbol{P_m} \rho^{(n)}(x) & \text{if } x \in S, \\ (\boldsymbol{I} - \beta \boldsymbol{P_m}) \rho^{(n)}(x) & \text{otherwise.} \end{cases} \tag{18}$$

A similar but not always equivalent way of describing this algorithm in a recursive form would be:

$$\rho^{(n+1)} = [\boldsymbol{P_s P_m} + \boldsymbol{P_{\underline{s}}}(\boldsymbol{I} - \beta \boldsymbol{P_m})] \rho^{(n)} \,, \tag{19}$$

with $\boldsymbol{P_{\underline{s}}} = (\boldsymbol{I} - \boldsymbol{P_s})$ the complement of the projector $\boldsymbol{P_s}$. Eqs. [12,14] can be used to describe the steps in terms of the gradients of the error metrics (see Sec. IV for details).

It is often used in conjunction of the ER algorithm, alternating several HIO iterations and one ER iteration (HIO(20)+ER(1) in our case). In particular one or more ER steps are used at the end of the iteration. Elser [39] pointed out that this may not be the best way of finding the final solution. The *iterate* ρ^n can converge to a *fixed point* ($\rho^{n+1} = \rho^n$), which may differ from the solution $\bar{\rho}$ ($\boldsymbol{P_s}\bar{\rho} = \boldsymbol{P_m}\bar{\rho} = \bar{\rho}$). However the solution $\bar{\rho}$ can be easily obtained from the fixed point:

$$\begin{aligned} \bar{\rho}_m^n &= \boldsymbol{P_m} \rho^n \,, \\ \bar{\rho}_s^n &= (1 + \tfrac{1}{\beta}) \boldsymbol{P_s P_m} \rho^n - \tfrac{1}{\beta} \boldsymbol{P_s} \rho^n \,, \end{aligned} \tag{20}$$

where $\bar{\rho}_m$ and $\bar{\rho}_s$ should coincide, or else their difference can be used as an error metric. See [39] for further details. *Difference Map* is a general set of algorithms, [39],

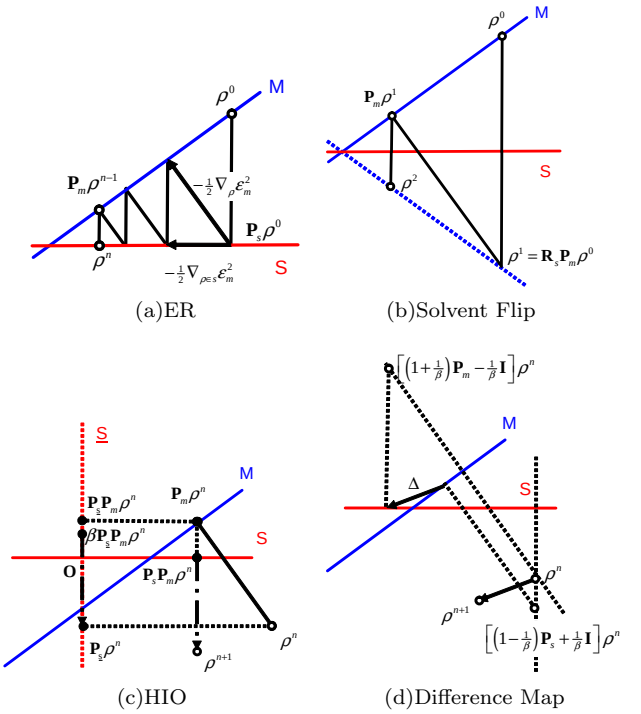


FIG. 3: Geometric representation of various algorithms using a simplified version of the constraint - two lines intersecting: (a) Error Reduction algorithm: we start from a point on the modulus constraint by assigning a random phase to the diffraction pattern. The projection onto the modulus constraint finds the point on the set which is nearest to the current one. The arrows indicate the gradients of the error metric. (b) The speed of convergence is increased by replacing the projector on the support with the reflector. The algorithm jumps between the modulus constraint (solid diagonal line) and its mirror image with respect to the support constraint (dotted line). (c) Hybrid input output, see text (Eq. 18). The space perpendicular to the support set is represented by the vertical dotted line $\underline{S}$.(d) Difference Map, see text (Eq. 21).

which requires 4 projections (two time-consuming modulus constraint projections) (Fig. 3(d)):

$$\begin{aligned} \rho^{(n+1)} = \{ \; & \boldsymbol{I} \; + \beta \boldsymbol{P_s} \, [(1+\gamma_s)\boldsymbol{P_m} - \gamma_s \boldsymbol{I}] \\ & - \; \beta \boldsymbol{P_m} \, [(1+\gamma_m)\boldsymbol{P_s} - \gamma_m \boldsymbol{I}]\} \rho^{(n)} \,, \end{aligned} \tag{21}$$

the solution corresponding to the fixed point is described in the same article [39]. We will use in the upcoming tests what the author suggested as the optimum with $\gamma_s = -\beta^{-1}$, $\gamma_m = \beta^{-1}$.

The *Averaged Successive Reflections* (ASR) [40] is:

$$\rho^{(n+1)} = \tfrac{1}{2}[\boldsymbol{R_s R_m} + \boldsymbol{I}] \rho^{(n)} \,. \tag{22}$$

The *Hybrid Projection Reflection* (HPR) [41] is derived from a relaxation of ASR:

$$\begin{aligned} \rho^{(n+1)} &= \tfrac{1}{2}[\boldsymbol{R_s}\,(\boldsymbol{R_m} + (\beta-1)\boldsymbol{P_m}) \\ &+ \; \boldsymbol{I} + (1-\beta)\boldsymbol{P_m}] \rho^{(n)} \,. \end{aligned} \tag{23}$$

It is equivalent to HIO if positivity (Sec. III A) is not enforced but it is written in a recursive form, instead of a case by case form such as Eq. 18. It is also equivalent to the Difference Map for $\gamma_1 = -1$, $\gamma_2 = \beta^{-1}$. Finally *Relaxed Averaged Alternating Reflectors* RAAR (previously named RASR) [42]

$$\rho^{(n+1)} = \left[\tfrac{1}{2}\beta\left(\boldsymbol{R_s R_m} + \boldsymbol{I}\right) + (1-\beta)\boldsymbol{P_m}\right]\rho^{(n)}. \quad (24)$$

For $\beta = 1$, HIO, HPR, ASR and RAAR coincide.

The first test is performed on the simplest possible case: find the intersection between two lines. Fig. 4 shows the behavior of the various algorithms, The two sets are represented by a horizontal blue line (support) and a tilted black line (modulus). ER simply projects back and forth between these two lines, and moves along the support line in the direction of the intersection. Solvent Flip projects onto the modulus, 'reflects' on the support, and moves along the reflection of the modulus constraint onto the support. The solvent flipping algorithm is slightly faster than ER due to the increase in the angle of the projections and reflections. HIO and variants (ASR, Difference Map, HPR and RAAR) move in a spiral around the intersection eventually reaching the intersection. For similar β RAAR behaves somewhere in between ER and HIO with a sharper spiral, reaching the solution much earlier. Alternating 20 iteration of HIO and 1 of ER (HIO(20)+ER(1)) considerably speeds up convergence.

When a gap is introduced between the two lines (Fig. 4(b)) so that the two lines do not intersect, HIO and variants move away from this local minimum in search for another 'attractor' or local minimum. This shows how these algorithms escape from local minima and explore the multidimensional space for other minima. ER, Solvent Flip, RAAR converge to or near the local minimum. By varying β RAAR becomes a local minimizer for small β, and becomes like HIO for $\beta \simeq 1$. ER, solvent flip HIO+ER converge to the local minimum in these tests.

A more realistic example is shown in Fig. 5. Here the circumference of two circles represent the modulus constraint, while the support constraint is represented by a line. The two circles are used to represent a non-convex set with a local minimum. It is difficult to represent a true modulus constraint in real space. For a representation of the modulus constraint in reciprocal space see [44]. The advantage of this example is the simplicity in the 'modulus' projector operator (it projects onto the closest circle). Although a real modulus constraint projector is not as simple as the one used in this example, there are similarities: each Fourier space point provides an n-dimensional quadratic equation.

We start from a position near the local minimum. ER, solvent flip and HIO+ER all fall into this trap (Fig. 5(a)), although increasing the interval between ER iterations in the HIO+ER algorithm would allow it to escape this local minimum. HIO and variants move away from the local minimum, 'find' the other circle, but converge to the center of the circle, with all but Diff. Map. not reaching a solution. In the center of the circle the projection on the modulus constraint becomes 'multivalued', and its distance metric is 'non-smooth'. The introduction of a small a random number added to the resulting solution at every step allows all the HIO-type codes to escape stagnation and find the solution (Fig. 5(b)). The random number can be as low as the numerical precision of the computer. For β reduced to 0.9, RAAR would not reach the solution, but converge close to the local minimum. As a latest test in this series Fig. 5(c), shows the behavior of the algorithms when the support is tangent to the circle, the two solutions coincide, and the the two constraints are parallel. The only algorithm to reach the solution is RAAR, but HIO+ER would also reach the solution if the interval between ER steps was sufficiently large. The other algorithms reach a fixed point from which the solution can be obtained (Eq. 20).

Another variant of the HIO algorithm was proposed by Takajo et al. [57] in which the Hybrid Input Output algorithm is used with an infinitesimally small β (Infinitesimal HIO, or IHIO). This approach has some important pedagogical advantages. IHIO contains an converging component which normally dominates, and a diverging component that has an effect only when a local minimum is found. Fig. 6 shows the behavior of IHIO compared to the standard HIO with $\beta = 0.75$. While HIO spirals around local minima, IHIO iterations move smoothly toward a local minimum, then move to the next local minimum. Both algorithms reach a *fixed point* of the iteration from which the solution can be found using Eq. 20, [39].

A. Positivity

The situation changes slightly when we consider the positivity constraint. The previous definitions of the algorithms still apply just replacing $\boldsymbol{P}_S$ with $\boldsymbol{P}_{S+}$:

$$\boldsymbol{P}_{S+} = \begin{cases} \rho(x) & \text{if } x \in S, \,\&\, \rho(x) \geq 0 \\ 0 & \text{otherwise.} \end{cases} \quad (25)$$

The only difference is HIO which becomes:

$$\rho^{(n+1)} = \begin{cases} \boldsymbol{P_m}\rho^{(n)}(x) & \text{if } x \in S \text{ and } \boldsymbol{P_m}\rho^{(n)}(x) \geq 0 \\ (1-\beta\boldsymbol{P_m})\rho^{(n)} & \text{otherwise.} \end{cases} \quad (26)$$

Fig. 7(a) shows that HIO bounces at the $x = 0$ axis. As the positivity constraint gets closer to the solution, none of the algorithms converges to the solution (Fig. 7(b)), with the HIO-type algorithms bouncing between the regions closer to the two circles. Only Difference Map for $\beta > 1$ converges (Fig. 7(c)). Also HIO+ER would reach the solution for larger intervals between ER iterations.

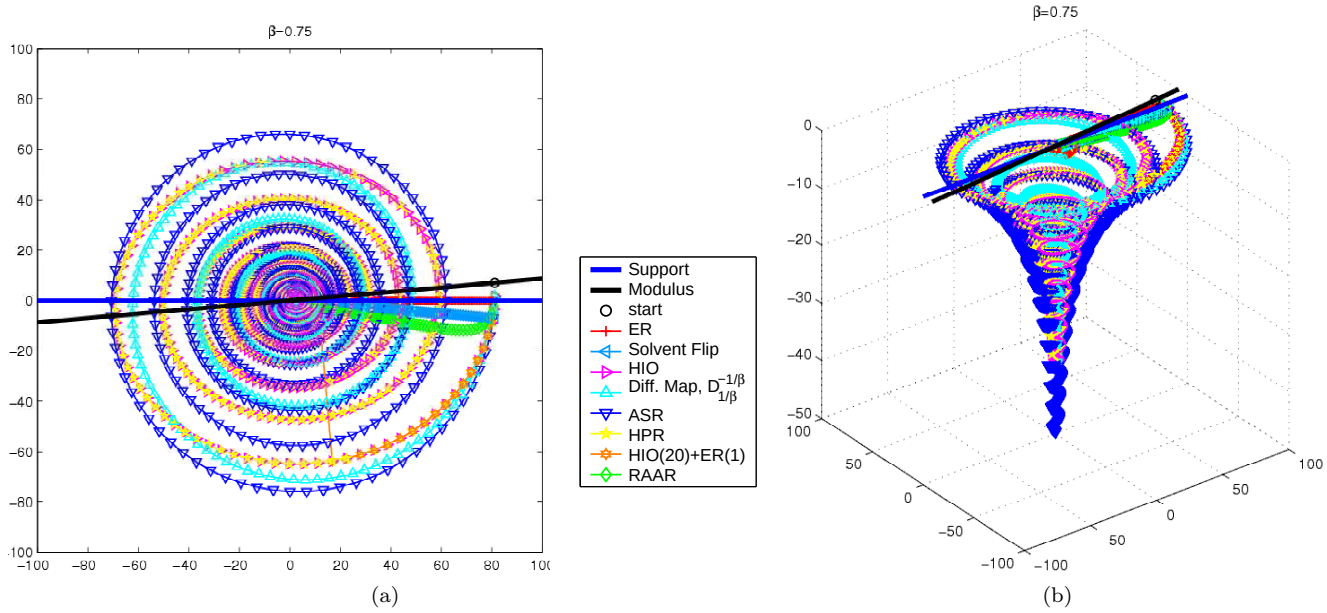


FIG. 4: The basic features of the iterative projection algorithms can understood by this simple model of two lines intersecting (4(a)). The aim is to find the intersection. The ER algorithm and the Solvent flipping algorithms converge in some gradient type fashion (the distance to the two sets never increases), with the solvent flip method being slightly faster when the angle between the two lines is small. HIO and variants move slightly in the direction where the gap between the two projections decreases, but at the same time in the direction of the gap, following a spiral path. When the two lines do not intersect (4(b), HIO and variants keep moving in the direction of the gap. ER, Solvent Flipping and RAAR converge at (or close to) the local minimum.

IV. STEEPEST DESCENT, CONJUGATE GRADIENT AND MIN-MAX ALGORITHMS

Conjugate gradient [56] methods (such as [19], and SPEDEN [43], which uses several other constraints) cannot be represented by simple geometrical models with one dimensional support sets used in the previous section. For these gradient based methods, we turn to a simple phase retrieval example. Takajo et al. [57] suggest using a simple image with 2 unknown pixels to reduce the phase retrieval problem to a 2-dimensional search. Fig. 8 shows the error metric ε_m as a function of the unknown variables (the two unknown pixel values), and the behavior of the ER algorithm toward the local minima. ER moves in the direction of the steepest descent, however the step length is not optimized to reach the local minimum in that direction since it is only one component of the full gradient (Fig. 8(a). In the standard steepest descent method, once the gradient of the function to be minimized is computed, one performs a line search of the local minimum in the descent direction.

$$\begin{aligned} &\min_\alpha \varepsilon_m^2\left(\rho+\alpha\Delta\rho\right) \\ &\Delta\rho = -\tfrac{1}{2}\nabla_s\varepsilon_m^2(\rho) = -\boldsymbol{P}_s[\boldsymbol{I}-\boldsymbol{P}_m]\rho, \end{aligned} \tag{27}$$

where $\nabla_s = \boldsymbol{P}_s\nabla_\rho$ is the gradient with respect to ρ_s. If any further step in the direction of the current step does not decrease the error metric, the gradient direction must be perpendicular to the current one. In other words the current step and the next step become orthogonal:

$$\begin{aligned} \tfrac{\partial}{\partial\alpha}\varepsilon_m^2(\rho+\alpha\Delta\rho) &= \langle\Delta\rho|\boldsymbol{P}_s[\boldsymbol{I}-\boldsymbol{P}_m]\left(\rho+\alpha\Delta\rho_s\right)\rangle_r\,, \\ 0 &= \langle\Delta\rho_s|[\boldsymbol{I}-\boldsymbol{P}_m]\left(\rho+\alpha\Delta\rho_s\right)\rangle_r\,. \end{aligned} \tag{28}$$

where $\langle\boldsymbol{x}|\boldsymbol{y}\rangle_r = \Re\,\boldsymbol{x}^\dagger\cdot\boldsymbol{y}$. The line search algorithm can use ϵ_m^2, and/or its derivative in Eq. 28. Evaluation of ε_m^2 and its derivative requires multiple modulus projections, so it is advantageous to use reciprocal space representations, where the modulus projector is a diagonal operator and is fast to compute, while the support projection requires two Fourier transforms:

$$\tilde{\boldsymbol{P}}_m\tilde\rho(k) = \frac{\tilde\rho(k)}{|\tilde\rho(k)|}\sqrt{I(k)}\,, \tag{29}$$

$$\tilde{\boldsymbol{P}}_s\tilde\rho(\boldsymbol{k}) = \mathcal{F}\boldsymbol{P}_s\mathcal{F}^{-1}\tilde\rho(\boldsymbol{k})\,, \tag{30}$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ represent the forward and inverse Fourier transforms respectively. At each iteration we add to the current image $\tilde\rho$, the step $\Delta\tilde\rho$:

$$\begin{aligned} \tilde\rho &\rightarrow \tilde\rho+\alpha\Delta\tilde\rho_s \\ \Delta\tilde\rho_s &= -\tilde{\boldsymbol{P}}_s[\boldsymbol{I}-\tilde{\boldsymbol{P}}_m]\tilde\rho\,, \end{aligned} \tag{31}$$

with α optimized to minimize the error metric (Fig. 8(b)).

The steepest descent method is known to be inefficient in the presence of long narrow valleys, where imposing that successive steps be perpendicular causes the algorithm to zig-zag down the valley. This problem is solved

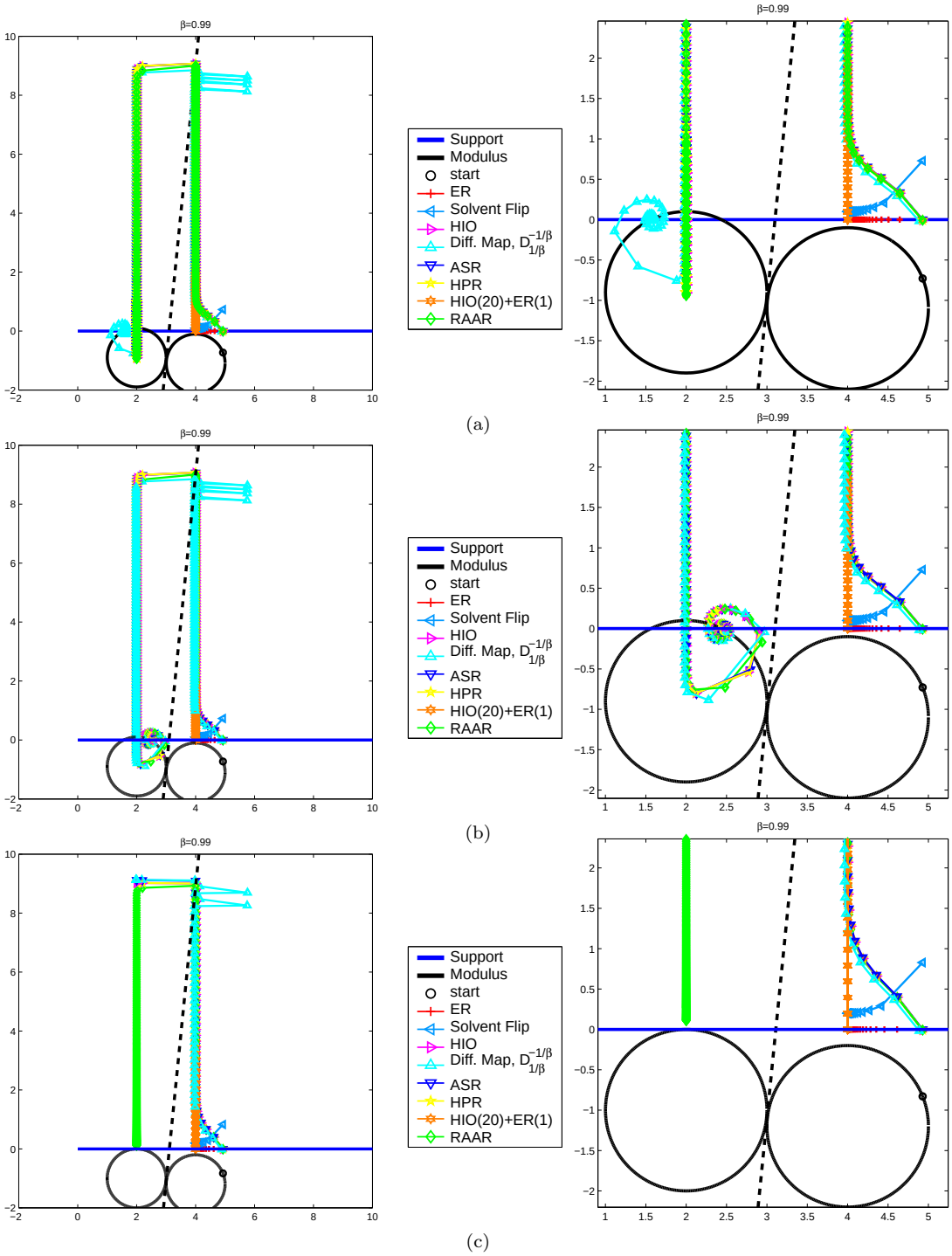


FIG. 5: The horizontal line represents a support constraint, while the two circles represent a non convex constraint, i.e. the modulus constraint. The dashed line divides the region closer to one circle from the other. The starting point is on the circle to the right, possessing a local minimum distance to the line. (a) The gradient-type (ER and Solvent Flip) algorithms converge to the local minimum, while HIO and variants move away from the local minimum in the direction of the gap (vertical) until they reach the region where the second circle is closer (delimited by the dashed line). From here they try to move in the same spiral-like path of the two lines (Fig. 4) until they reach the point where the projection on the circle and the line are parallel, and start moving toward the the center of the circle which has the correct solution. They stagnate in the center of the circle where the projection is multivalued. Only the Diff. Map reaches one of the two solutions. The addition of a small value of the order of the numerical precision after each iteration solves this stagnation (b). When one of the circles just touches the other constraint most algorithms either get stuck near the local minimum or stagnate. RAAR is the only one that reaches the vicinity of the solution (c).

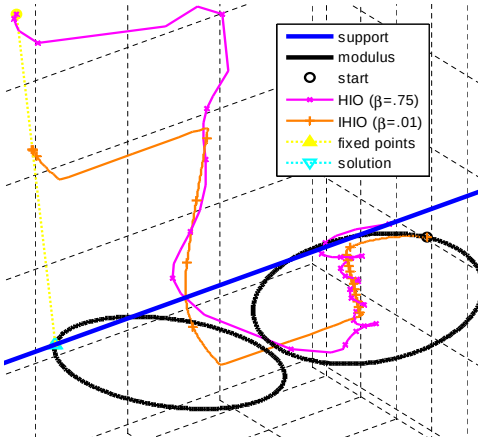


FIG. 6: Infinitesimal HIO and standard HIO in the case of two tilted circles (modulus) with one intersecting a straight line (support) in one point solution. The gridlines represent the three dimensional axes in the object space. Both algorithm jump through the same local minima, reaching the same fixed points. The corresponding solution obtained by Eq. 20 represents the only intersection of the two constraints. While HIO spirals around local minima, IHIO iterations move smoothly toward a local minimum, then move to the next local minimum.

by the nonlinear conjugate gradient method [55]. Instead of moving in the direction of steepest descent $\Delta\rho_s$, we move in the direction $\Lambda\rho_s$:

$$\Lambda\rho_s^{(n)} = \begin{cases} \Delta\rho_s^{(n)} & \text{if n=1} \\ \Delta\rho_s^{(n)} + \gamma_s \Lambda\rho_s^{(n-1)} & \text{otherwise.} \end{cases} \quad (32)$$

with

$$\gamma_s = \max\left\{ \frac{\left\langle \Delta\rho_s^{(n)} | \Delta\rho_s^{(n)} - \Delta\rho_s^{(n-1)} \right\rangle_r}{\|\Delta\rho_s^{(n-1)}\|^2}, 0 \right\} . \quad (33)$$

The presence of local minima shown in the previous chapters however will cause stagnation of steepest and conjugate gradient methods, preventing global convergence (Fig. 8(c)).

A. Feedback and the saddle point problem

HIO and variants are able to escape local minima and reach the solution faster (Fig. 8(d)). However as in ER, the step length is not optimized, the algorithm keeps moving in the same direction for several steps. Combining the ideas of the conjugate gradient or the steepest descent methods and HIO could considerably speed-up convergence. As discussed earlier, the component of the iteration step $\Delta\tilde{\rho}$ inside the support $\Delta\tilde{\rho}_s = \tilde{\boldsymbol{P}}_s\Delta\rho$ descends along the $\tilde{\rho}_s$ components of the gradient of the error metric $\varepsilon_m^2(\tilde{\rho})$:

$$\Delta\tilde{\rho}_s = -\nabla_s^\dagger \varepsilon_m^2(\tilde{\rho}) = \tilde{\boldsymbol{P}}_s[\tilde{\boldsymbol{P}}_m - \boldsymbol{I}]\rho\,, \quad (34)$$

where $\tilde{\rho}$ and its conjugate $\tilde{\rho}^\dagger$ are treated as independent variables and $\nabla_s^\dagger = \tilde{\boldsymbol{P}}_s\nabla^\dagger$ is the gradient with respect to $\tilde{\rho}_s^\dagger$. The feedback component $\beta\Delta\tilde{\rho}_{\underline{s}}$ constraints the recovered amplitudes $|\tilde{\rho}|$ by minimizing the function $\varphi(\tilde{\rho}) = 2\sum\sqrt{I}|\tilde{\rho}|$ along the $\tilde{\rho}_{\underline{s}} = \tilde{\boldsymbol{P}}_{\underline{s}}\tilde{\rho}$ components:

$$\Delta\tilde{\rho}_{\underline{s}} = -\nabla_{\underline{s}}^\dagger\varphi(\tilde{\rho}) = -\tilde{\boldsymbol{P}}_{\underline{s}}\tilde{\boldsymbol{P}}_m\tilde{\rho} \quad (35)$$

Integrating the gradients along a simple path, e.g. from 0 to $\tilde{\rho}_s$ and from $\tilde{\rho}_s$ to $\tilde{\rho}_s + \tilde{\rho}_{\underline{s}}$ provides the Lagrangian function $\mathcal{L}$:

$$\mathcal{L}(\tilde{\rho}_s, \tilde{\rho}_{\underline{s}}) = \sum \left| [\tilde{\boldsymbol{P}}_m - \boldsymbol{I}]\tilde{\rho}_s \right|^2 - 2\sqrt{I}\left\{ \left|\tilde{\rho}_s + \tilde{\rho}_{\underline{s}}\right| - |\tilde{\rho}_s| \right\} \quad (36)$$

HIO/HPR/ASR algorithms move toward the saddle point of such Lagrangian, the minimum in the direction $\tilde{\rho}_s$ ($\nabla_s^\dagger\mathcal{L} = -\Delta\tilde{\rho}_s$) and the maximum in the direction $\tilde{\rho}_{\underline{s}}$ ($\nabla_{\underline{s}}^\dagger\mathcal{L} = \Delta\tilde{\rho}_{\underline{s}}$). They do so using a gradient optimization strategy, where the step is proportional to the gradient but with one sign reversal, moving toward the saddle point, rather than in the descent direction. This minmax saddle point problem is common in fields as various as game theory, economics, physics, and engineering [??]. Minimization problems are helped by the fact that we want the function to get smaller. The saddle can be higher or lower than the current value, although the direction toward the saddle is still indicated by the gradient. The optimization of the parameter α ($\beta = \alpha$) is obtained by increasing α until the current and next search directions become perpendicular to one another (Fig. 8(e)):

$$\left\langle \Delta\tilde{\rho} | \{\tilde{\boldsymbol{P}}_s[\tilde{\boldsymbol{P}}_m - \boldsymbol{I}] - \tilde{\boldsymbol{P}}_{\underline{s}}\tilde{\boldsymbol{P}}_m\}(\tilde{\rho} + \alpha\Delta\tilde{\rho}) \right\rangle = 0 \quad (37)$$

which can be computed avoiding time consuming Fourier transforms (required by the projectors $\tilde{\boldsymbol{P}}_s, \tilde{\boldsymbol{P}}_{\underline{s}}$), using $\langle\Delta\tilde{\rho}|\tilde{\boldsymbol{P}}_s x + \tilde{\boldsymbol{P}}_{\underline{s}} y\rangle = \langle\Delta\tilde{\rho}_s|x\rangle + \langle\Delta\tilde{\rho}_{\underline{s}}|y\rangle$. In analogy to the the conjugate gradient method, one could substitute the search direction $\Delta\tilde{\rho}$ with $\Lambda\tilde{\rho}$ as in Eq. 32 (Fig. 8(f)). Further improvement should be obtained by replacing the one dimensional search with a two dimensional optimization of the saddle point (Fig. 8(g)):

$$\min_\alpha \max_\beta \mathcal{L}(\tilde{\rho}_s + \alpha\Delta\tilde{\rho}_s, \tilde{\rho}_{\underline{s}} + \beta\Delta\tilde{\rho}_{\underline{s}}), \quad (38)$$

increasing α, β in the direction $(-\frac{\partial}{\partial\alpha}, \frac{\partial}{\partial\beta})\mathcal{L}$, until both components are equal to 0:

$$\begin{cases} \left\langle \Delta\tilde{\rho}_s | \tilde{\boldsymbol{P}}_s[\tilde{\boldsymbol{P}}_m - \boldsymbol{I}]\left(\tilde{\rho} + \alpha\Delta\tilde{\rho}_s + \beta\Delta\tilde{\rho}_{\underline{s}}\right) \right\rangle_r = 0 \\ \left\langle \Delta\tilde{\rho}_{\underline{s}} | - \tilde{\boldsymbol{P}}_{\underline{s}}\tilde{\boldsymbol{P}}_m\left(\tilde{\rho} + \alpha\Delta\tilde{\rho}_s + \beta\Delta\tilde{\rho}_{\underline{s}}\right) \right\rangle_r = 0. \end{cases} \quad (39)$$

i.e. when the components $\Delta\tilde{\rho}_s$ and $\Delta\tilde{\rho}_{\underline{s}}$ of successive steps become perpendicular. In the conjugate gradient version, one can use a common γ value to obtain the conjugate direction (Eq. 33) Fig. 8(h), or calculate $\gamma_s, \gamma_{\underline{s}}$ independently 8(i).

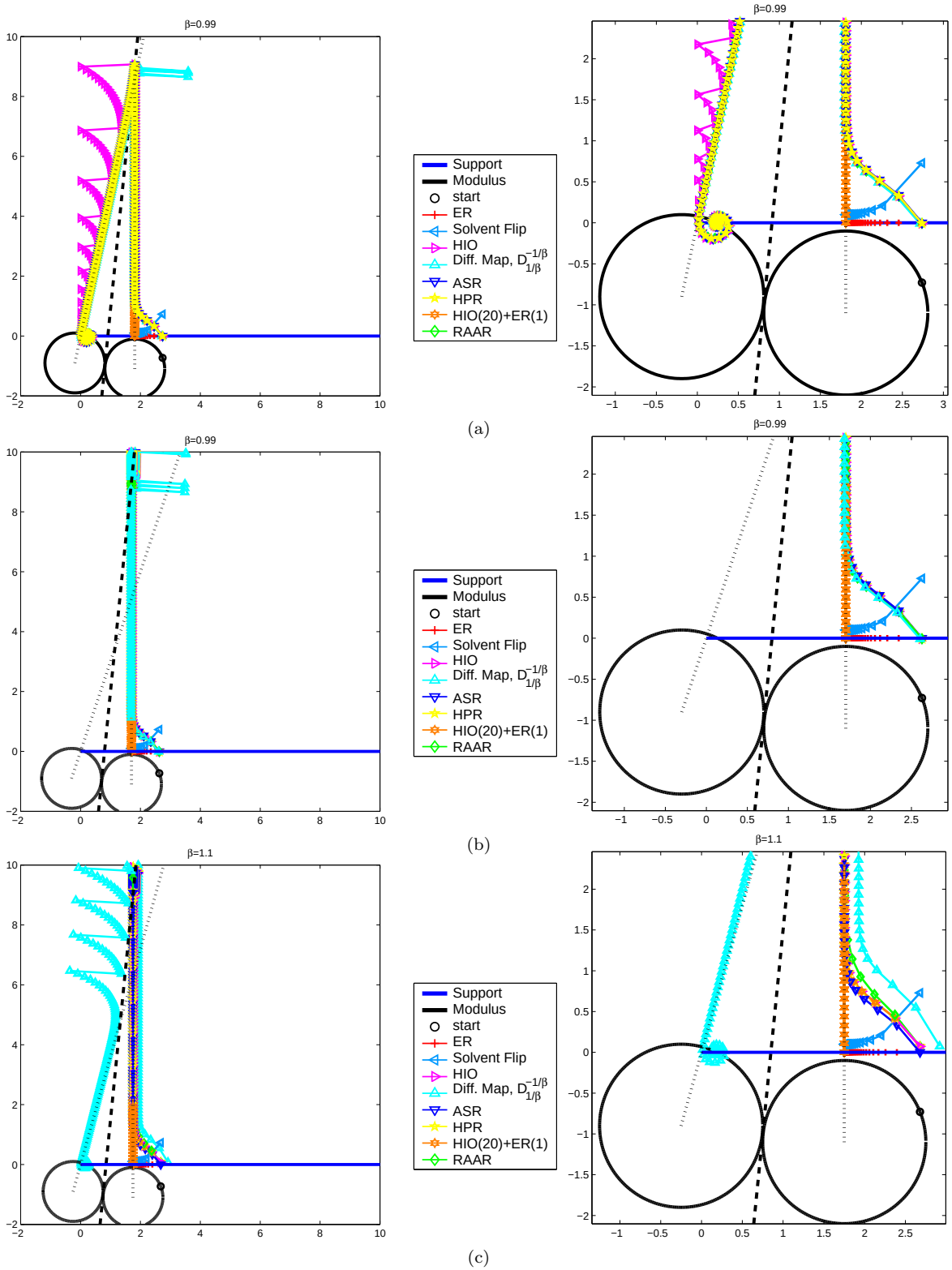
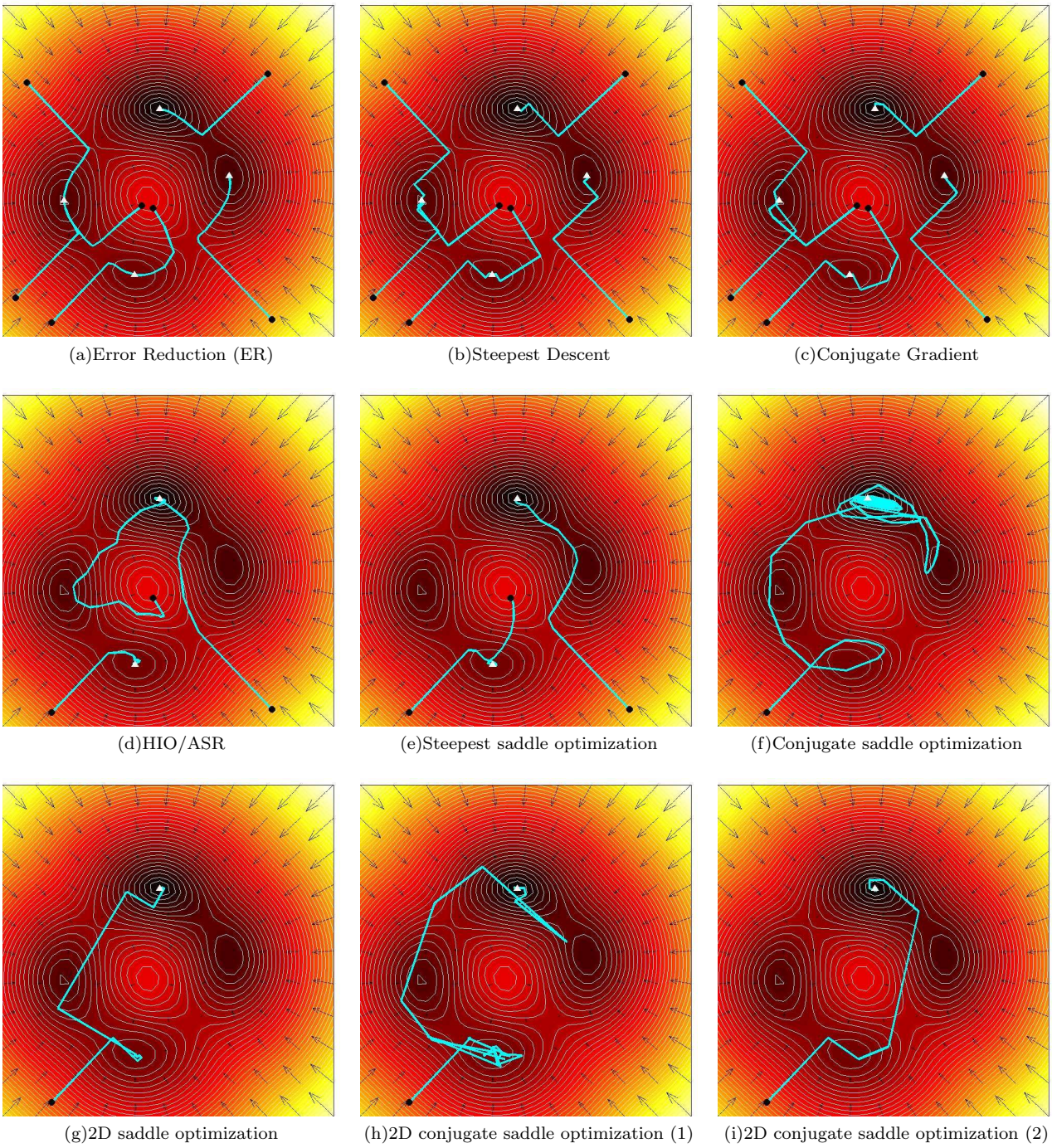


FIG. 7: This figure shows the behavior of the algorithms when the positivity constraint is introduced, with the support constraint represented by a horizontal line originating from the 0 ($x \geq 0$). (a) The starting point is again on the circle to the right, close to a local minimum. HIO and variant move away from the local minimum in the direction of the gap until they reach the region where the circle to the left is closer. Instead of moving in a spiral like fashion, the iterations move close to the line joining the center of the left circle to the origin (represented by a dotted line), except for HIO that bounces on the x=0 axis. (b) The solution is very close to 0, and the dotted line becomes more tilted. The various algorithms after moving in the vertical direction away from the local minimum, reach the dashed line and start moving toward the tilted dotted line, falling back in the region closer to the first minimum. These algorithms bounce between the regions closer to each circle without reaching the solution. With $\beta > 1$, i.e. inverting the order of the operators, Diff. Map converges, and RAAR diverges, while HIO HPR and ASR stagnate.



(a)Error Reduction (ER) (b)Steepest Descent (c)Conjugate Gradient

(d)HIO/ASR (e)Steepest saddle optimization (f)Conjugate saddle optimization

(g)2D saddle optimization (h)2D conjugate saddle optimization (1) (i)2D conjugate saddle optimization (2)

FIG. 8: A simple 2-D phase retrieval problem: only two variables are unknown, while the others are known. The solution -the global minimum- is the top minimum in the figures. The colormap and contour lines represents the error metric $\varepsilon_m(\rho_s)$, and the descent direction is indicated by the arrows. The Error Reduction algorithm (a) proceeds toward the local minimum without optimizing the step length and stagnates at the local minima. The steepest descent method (b) moves toward the local minimum with a zig-zag trajectory, while the conjugate gradient method reaches the solution faster (c). The Hybrid Input Output method avoids most stagnations, however with some rare unlucky initial guess stagnates in a local minimum (d). The saddle point optimization moves with larger steps but stagnates in the same local minimum as HIO (e). The conjugate gradient version avoids stagnation(f). The saddle point optimization using a two dimensional search of the saddle point does not stop at the local minimum but in the "steepest" version the algorithm spends several iterations to escape the local minimum (g). The conjugate gradient version (h, i) reaches the solution faster if the conjugate direction is chosen independently (i).

V. CONCLUSIONS

Recombining numerically rather than optically light scattered by illuminated objects is being increasingly useful at tackling unsolved scientific problems. The new instruments replacing lenses are the iterative projection algorithms for phase retrieval. We reviewed the algorithms from the literature, and evaluated them with geometrical sets and simple phase retrieval problem. ER is a simple but powerful local minimizer using a gradient search of the minimum of the error metric. Its rate of convergence can be increased optimizing the step length along the descent direction and using either the steepest descent or conjugate gradient methods (such as SPEDEN [43]) reviewed in section IV. HIO and variants are very powerful in escaping local minima, but in several situations fail to converge. When positivity is introduced, the recursive version of HIO (HPR) converges more 'smoothly' to the solution without bouncing on the $x = 0$ axis. Alternating between HIO and ER with the correct intervals would have worked in all the examples shown above, however the correct interval is not known in advance. RAAR is a good (single parameter) way to change from 'global' to local minimizer. Difference Map is successful in a few more of the examples shown above for the proper choice of β, however it involves 2 time consuming modulus constraint operations. The connection between the feedback and the saddle point optimization is discussed in section IV. Algorithms with feedback (HIO, HPR and ASR) correspond to a gradient search of the saddle point, and acceleration strategies such as conjugate gradient method for the minmax problem have been discussed.

The Solvent flipping algorithm does not show much success in the examples shown above. Despite this it was used to improve images [49], and in a modified form to solve 3D structures ab-initio [53].

Acknowledgments

This work was performed under the auspices of the U.S. Department of Energy by the Lawrence Livermore National Laboratory under Contract No. W-7405-ENG-48 and the Director, Office of Energy Research. This work was funded by the National Science Foundation through the Center for Biophotonics. The Center for Biophotonics, a National Science Foundation Science and Technology Center, is managed by the University of California, Davis, under Cooperative Agreement No. PHY0120999. D. R. Luke provided very useful comments.

[1] M. R. Howells, et al. J. Elect. Spect. and Rel. Phen. (2004), (arXiv:physics/0502059).
[2] J. C. Solem, G. C. Baldwin, Science 218, 229-235 (1982).
[3] R. Neutze et al, Nature **406**, (2000) 752.
[4] J. C. H. Spence, R. B. Doak, Phys. Rev. Lett. **92**, (2004) 198102.
[5] R. Gerchberg and W. Saxton, Optik **35**, 237 (1972).
[6] J. R. Fienup, Opt. Lett. **3**, (1978) 27.
[7] J.R. Fienup, J.C. Marron, T.J. Schulz and J.H. Seldin, Appl. Opt. **32**, (1993) 1747-1768.
[8] J. C. H. Spence, U. Weierstall , M. Howells Phil. Trans. A 360(1974), (2002), 875-895.
[9] J.M. Zuo, I. Vartanyants, M. Gao, R. Zhang and L.A. Nagahara, Science 300, (2003) 1419-1421.
[10] J. Wu, U. Weierstall, J. C. H. Spence, Nature Materials, 4, (2005), 912.
[11] R. P. Millane, J. Opt. Soc. Am. A **7**, 394 (1990).
[12] International Tables for Crystallography Vol. F, Ed. M. G. Rossmann and E. Arnold, The International Union of Crystallography, Kluwer Academic Publishers, Dordrecht/Boston/London 2001.
[13] R. H. T. Bates, Optik **61**, (1982) 247.
[14] D. Sayre, Acta Cryst. **5**, (1952) 843.
[15] D. Sayre, "Prospects for long-wavelength x-ray microscopy and diffraction", in Imaging Processes and Coherence in Physics, Schlenker, M., M. Fink, J. P. Goedgebuer, C. Malgrange, J. C. Viénot, R. H. Wade, (Ed), Lecture Notes in Physics, Vol. 112, 229-235, Springer-Verlag, Berlin, 1980.
[16] J. Miao, D. Sayre, and H. N. Chapman, J. Opt. Soc. Am A **15** (1998) 1662.
[17] W. McBride, N. L. O'Leary, and L. J. Allen, Phys. Rev. Lett. 93, 233902 (2004).
[18] J. R. Fienup, Opt. Eng. **19**, (1980) 297.
[19] J. R. Fienup, Appl. Opt. **21**, (1982) 2758.
[20] J. Miao, P. Charalambous, J. Kirz, D. Sayre, Nature **400**, (1999) 342.
[21] G. J. Williams, M. A. Pfeifer, I. A. Vartanyants, I. K. Robinson, Phys. Rev. Lett. **90**, (2003) 175501.
[22] J. Miao, et al. Phys. Rev. Lett., **89** (2002), 088303.
[23] J. W. Miao, K. O. Hodgson, T. Ishikawa, C. A. Larabell, M. A. LeGros, Y. Nishino, Proc. Nat. Ac. Sci. **100**, (2003) 110.
[24] T. Beetz, et al. Nucl. Instrum. Meth. A **545**, (2005) 459.
[25] D. Shapiro, P. Thibault, T. Beetz, V. Elser, M. Howells, C. Jacobsen, J. Kirz, E. lima, H. Miao, A. Neiman, D. Sayre, PNAS 102 (43), (2005) 1543-1546.
[26] E. Lima, et al. XRM proceedings (2005).
[27] M. R. Howells et al, Proc. SPIE **4783**, (2002) 65.
[28] H. N. Chapman, A. Barty, S. Marchesini, A. Noy, C. Cui, M. R. Howells, R. Rosen, H. He, J. C. H. Spence, U. Weierstall, T. Beetz, C. Jacobsen, D. Shapiro, J. Opt. Soc. Am. A , in press (arXiv:physics/0509066).
[29] A. Barty et al. in preparation
[30] S. Marchesini et al. Phys. Rev. B **68**, (2003) 140101(R), (arXiv:physics/0306174).
[31] J. C. H. Spence, U. Weierstall, and M. R. Howells, Ultramicros. 101, (2004) 149.
[32] J. W. Goodman, "Statistical Optics" New York: Wiley (1985).
[33] R. Barakat and G. Newsam, "Necessary conditions for a unique solution to two-dimensional phase recovery," J.

Math. Phys. 25, 3190-3193 (1984).
[34] M. J. Buerger, "Vector space and its application in crystal structure investigation", Wiley, New York (1959).
[35] S. Marchesini, H. N. Chapman, A. Barty, M. R. Howells, J. C. H. Spence, C. Cui, U. Weierstall, and A. M. Minor, to be published in the XRM 2005 proceedings, Jap. J. Appl. Phys. (arXiv:physics/0510033).
[36] J. N. Cederquist, J. R. Fienup, J. C. Marron, R. G. Paxman, Opt. Lett. **13**, 619. (1988).
[37] A. Levi and H. Stark, J. Opt. Soc. Am. A **1**, 932-943 (1984).
[38] H. Stark, *Image Recovery: Theory and applications*. (Academic Press, New York, 1987).
[39] V. Elser, J. Opt. Soc. Am. A **20**, 40 (2003).
[40] H. H. Bauschke, P. L. Combettes, and D. R. Luke. J. Opt. Soc. Am. A **19**, 1334-1345 (2002).
[41] H. H. Bauschke, P. L. Combettes, and D. R. Luke, J. Opt. Soc. Am. A **20**, 1025-1034 (2003).
[42] D. R. Luke, Inverse Problems 21:37-50(2005), (arXiv:math.OC/0405208).
[43] S. P. Hau-Riege, H. Szöke, H. N. Chapman et al. (2004) (arXiv:physics.optics/0403091)
[44] D. R. Luke, J. V. Burke, R. G. Lyon, SIAM Review **44** 169-224 (2002).
[45] J. V. Burke and D. R. Luke. SIAM J. Control Opt. **42**, 576-595 (2003).
[46] P-T. Chen, M. A. Fiddy, C-W. Liao and D. A. Pommet, J. Opt. Soc. Am. A 13, (1996), 1524-31.
[47] J. R. Fienup, C. C. Wackerman, J. Opt. Soc. Am. A 3, 1897-1907 (Nov. 1986).
[48] L. M. Brègman, Sov. Math. Dokl. **6**, 688-692 (1965).
[49] J. P. Abrahams, A. W. G. Leslie, Acta Cryst. **D52**, 30-42 (1996)
[50] S. Marchesini et al. Optics Express **11**, (2003) 2344, (arXiv:physics/0308064).
[51] J. R. Fienup, A. M. Kowalczyk J. Opt. Soc. Am. A **7**, 450 (1990).
[52] V. Elser, Acta Cryst. **A59**, 201-209 (2003), (arXiv:cond-mat/0209690).
[53] G. Oszlányi and A. Süto, Acta Cryst. **A60**, 134-141 (2004) (arXiv:cond-mat/0308129).
[54] B. Carrozzini, G. L. Cascarano, L.De Caro, et al. Acta Cryst. A60, 331-338 (2004), (arXiv:physics.optics/0404073).
[55] W. H. Press, S. A. Teukolsky, W. T. Vetterling, Brian P. Flannery, Numerical Recipes in C, Cambridge University Press,
[56] M.R. Hestenes, Conjugate Direction Methods in Optimization, Springer-Verlag, New York, 1980
[57] H. Takayo, T. Takahashi, R. Ueda, M. Taninaka J. Opt. Soc. Am. A 15, 2849-61 (1998)

This work was performed under the auspices of the U. S. Department of Energy by University of California, Lawrence Livermore National Laboratory under contract W-7405-Eng-48.