Phase retrieval with Fourier-weighted projections

Manuel Guizar-Sicairos and James R. Fienup*

The Institute of Optics, University of Rochester, Rochester, New York, 14627, USA
**Corresponding author: fienup@optics.rochester.edu*

Received October 9, 2007; accepted December 16, 2007;
posted January 11, 2008 (Doc. ID 88391); published February 14, 2008

In coherent lensless imaging, the presence of image sidelobes, which arise as a natural consequence of the finite nature of the detector array, was early recognized as a convergence issue for phase retrieval algorithms that rely on an object support constraint. To mitigate the problem of truncated far-field measurement, a controlled analytic continuation by means of an iterative transform algorithm with weighted projections is proposed and tested. This approach avoids the use of sidelobe reduction windows and achieves full-resolution reconstructions. © 2008 Optical Society of America

OCIS codes: 100.5070, 100.3020, 100.6640, 260.1960, 340.7440.

1. INTRODUCTION

Coherent lensless imaging by phase retrieval, sometimes referred to as coherent diffractive imaging (CDI), is an active imaging approach in which an image of a target can be formed without the use of any imaging optics [1–4]. The object of interest is illuminated by a coherent beam, and the backscattered radiation or, alternatively, the radiation transmitted by a semitransparent object, propagates and is collected by an ordinary intensity detector array. The measured far-field intensity pattern $|F(\mathbf{u})|^2$, which is a speckle pattern for the case of a random-phase object, can then be processed to yield an image of the target $f(\mathbf{x})$, where a phase retrieval algorithm substitutes for optics in the image formation process, as illustrated in Fig. 1.

Hardware implementation of CDI is substantially simpler than other lensless imaging techniques, such as digital holography. It does not require a retro-reflector or glint that is well separated from the object nor a sharp corner or slit as required by reconstruction with a boundary-wave reference [5,6], nor are a mirror and a beam splitter needed as for a holographic reconstruction [7], nor is a frequency-stabilized local oscillator like that required for heterodyne detection needed. Furthermore, the illumination beam needs to be spatially coherent only over the width of the object.

In the proposed configuration, a diffraction-limited image can be formed with a resolution equivalent to what would be obtained by an optical system with an entrance pupil of the same size and shape as the detector array. For conventional imaging with a lens or mirror, to achieve finer resolution, both the diameter of the aperture and the axial length of the system must increase, and the cost and difficulty of creating a low-aberration system increases as well. CDI offers an alternative that can lead to high-resolution images with a thin, lightweight measurement system, the axial thickness of which need not increase with increasing diameter. Alternatively, this approach can be used to reconstruct images when aberration-compensated optical elements are not available for the desired wavelength, as is the case of x-ray diffraction imaging [8].

The detected, nonimaged intensity pattern should ideally be Nyquist sampled for intensity. This lateral sampling requirement is only half of that needed for an off-axis holographic reconstruction, for which one wishes to avoid the overlap of the reconstructed image with the autocorrelation term. For this imaging approach to work in reflection, the coherence length of the illuminating beam should be at least twice as long as the object is deep.

The Fourier transform relation is

$$F(\mathbf{u}) = |F(\mathbf{u})|\exp[i\phi(\mathbf{u})] = \int f(\mathbf{x})\exp(-i2\pi\mathbf{u}\cdot\mathbf{x})\mathrm{d}\mathbf{x}, \quad (1)$$

where $\mathbf{x}$ and $\mathbf{u}$ are the coordinates in object and Fourier domains, respectively. The phase retrieval problem considered here is the reconstruction of a complex-valued image $f(\mathbf{x})$ [or, equivalently, the retrieval of the phase distribution $\phi(\mathbf{u})$ in Fourier space], given the magnitude $|F(\mathbf{u})|$ of its Fourier transform and some knowledge of the object support S, where S represents a set of points outside of which the object is known to be zero. Throughout this paper, uppercase letters represent the Fourier transforms of their lowercase counterparts. Because the inverse Fourier transform of the intensity $|F(\mathbf{u})|^2$ is the image autocorrelation, this problem can be alternatively stated as reconstructing the image from its autocorrelation.

A common technique to solve the image-reconstruction phase retrieval problem is the use of iterative transform algorithms (ITAs), which involve transforming back and forth between object and Fourier spaces to find a Fourier pair in agreement with both the measured Fourier magnitude (the square root of the measured intensity) and the support in the object domain.

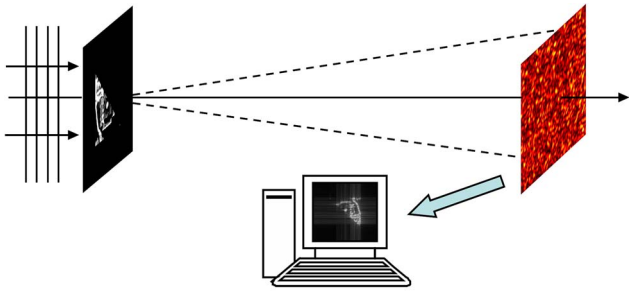

Fig. 1. (Color online) Setup for coherent lensless imaging by phase retrieval.

It is important to note that, because in CDI the object is coherently illuminated, the desired image $f(\mathbf{x})$ is often complex valued and a nonnegativity constraint cannot be used. Thus the phase retrieval problem becomes much more difficult [1,2,9–11], and the use of a technique that is robust and capable of escaping local minima, e.g., the hybrid input–output (HIO) algorithm [12], becomes crucial.

When developing or extending phase retrieval algorithms, it is a common practice to perform numerical tests to asses their performance. These numerical simulations of far-field diffraction patterns provide a controlled environment in which the actual solution is known and the effects of noise, loose support, or other issues can be studied in detail. The forward model to simulate the measured data may involve taking an image with random phase statistics, setting all the values outside of a given object support to zero, computing the fast Fourier transform (FFT), taking the square modulus, and using this as the far-field measured intensity pattern [1]. A problem that arises from this numerical simulation is that the Fourier data directly computed from an FFT of the finite-support object is aliased and presents nonphysical continuous wrap-around effects across the edge of the computational window. Because of this, using the entire Fourier array for reconstruction can yield an image that does not have any sidelobes.

In the real world, a coherently illuminated object will naturally backscatter energy that falls outside the detector array, thus making the field at the detector discontinuous at the edge of the computational window. For the case of an object with a strong phase, such as a diffusely scattering random-phase object, the field at the edge of the detector array is on average as bright as in the center of the detector, causing the discontinuity to be large. In contrast, the object in the numerical simulation is sampled, so the energy that would fall out of the detector wraps around because of aliasing and reappears inside the computational window. The values at one edge of the FFT are then continuous with the values at the opposite edge. The use of this simulated wrap-around intensity pattern for image reconstruction is advantageous for phase retrieval algorithms, because this far-field data is continuous in wrap-around space and is perfectly consistent with a finite support. This is, however, a nonphysical effect that might lead us to think that algorithms perform better than they would in a real application.

In the real world, even if the object of interest has a hard-edge support, any image of it will have sidelobes that extend beyond this support due to the finite aperture of the optical system. This applies to CDI since it reconstructs an image from a truncated measurement of backscattered radiation intensity. It should be pointed out that there is a subtle difference between the images that are computationally formed by phase retrieval and the equivalent physical optical system: while the image that is formed optically has sidelobes that extend indefinitely, the reconstructed image sidelobes wrap around due to sampling in the Fourier domain, thereby causing aliasing of the sidelobes in the image.

Because any image with a finite object support necessarily has a spectrum that extends indefinitely, the ITA is presented with an ill-posed (inconsistent) problem as it attempts to reconstruct an image of an object with a finite support from a truncated (finite) Fourier-domain measurement. Moreover, image sidelobes affect the algorithm performance because the ITA tries to set to zero the energy outside the support while the actual solution has some energy in that region due to the sidelobes. So the algorithms can drift away from the ideal reconstruction even if the true solution is used as starting guess.

For a realistic simulation of measured data, it is important to extract a portion (e.g., a square of half the size of the total array in each dimension) of the computed far-field intensity after the support constraint has been applied in object space and use only this portion for phase retrieval. This makes the simulated data more realistic by giving Fourier data without wrap around at the edges and including image sidelobes. It also substantially reduces the success rate of phase retrieval algorithms.

The presence of these image sidelobes was recognized early as a source of convergence problems for phase retrieval algorithms [4,10,11,13]. A technique used to circumvent this problem is to apply an apodization filter in the Fourier domain, e.g., multiply by a separable Hanning weighting function, to reduce the image sidelobes and allow successful reconstructions. This is accomplished, however, at the cost of reduced resolution in the resulting image.

In this paper, we propose a controlled analytic continuation in Fourier space to mitigate the problem of truncated measurement arising from the finite diameter of the detector array in CDI. By applying a weighted projection in the Fourier domain, we slightly extrapolate the measured data outside of the computational window. In this way we allow sidelobes to form around our far-field measured data and alleviate the inconsistency of trying to have hard edges in both domains. This is done without sacrificing resolution as when using Fourier apodization filtering. Use of a weighting function outside of the measurement area that varies smoothly has shown performance superior to a previously reported iterative extrapolation technique [14].

It should be stressed that, although this discussion is centered on hard-edge objects, for which these effects are most noticeable, the matter of concern is the nonphysical continuity at the edge of the Fourier computational window. So to achieve optimal results, it is important to include the effects of truncated measurement even when the object has smooth or tapered edges, as occurs, for example, in the case of an extended object with an illumination pattern constraint [4].

2. PHASE RETRIEVAL FOR COMPLEX-VALUED IMAGES

A far-field measured speckle intensity pattern was numerically simulated to test the proposed phase retrieval algorithm. A 256×256 complex-valued synthetic-aperture radar image of Michigan Stadium [15], shown in Fig. 2(a), was multiplied by the triangular binary object support shown in Fig. 2(b) to yield the finite-support object shown in Fig. 2(c). Making the support less than half the width and height of the complete array ensures that the Fourier intensity pattern will be Nyquist sampled. The shape of the support was chosen to be noncentrosymmetric to avoid the twin image problem and make the reconstruction easier [13]. The finite-support object was then Fourier transformed, as shown in Fig. 2(d), and the 128×128 central portion of the magnitude, shown in Fig. 2(e), was used for the simulated data collected by the detector array.

A diffraction-limited image computed from the 128×128 $F(\mathbf{u})$, the ideal solution to the phase retrieval problem, is shown in Fig. 2(f). Notice that this image has sidelobes that extend outside of the object support. Due to the large dynamic range of the image, the square root of the image magnitude is shown throughout this paper. For better detail visualization, the 128×128 images are upsampled by $2\times$ by embedding the 128×128 complex Fourier transform in a 256×256 array of zeros and computing the inverse Fourier transform. Only a 160×160 portion of the upsampled image is displayed.

A hard-edge support can be obtained by flood illuminating an object that is spatially finite and defines its own support. This is consistent with the problem of x-ray imaging of small structures or the observation of an airborne or spaceborne target. A hard-edge support can also be obtained by inserting an aperture mask in the plane of the object. The support can be estimated by using a low-resolution image of the object, for example by means of either an electronic microscope or a small telescope [10]. Alternatively, an upper bound on the support can be obtained directly from the support of the image autocorrelation [16]. In this approach a set that includes all possible object supports that can give rise to the autocorrelation support (excluding their reflections through the origin) can be obtained. This results in a support constraint that may be too loose to achieve a successful reconstruction of a complex-valued image. Algorithms that adaptively improve the support constraint as the iterations progress have been proposed [17]. Throughout this paper we will use an *a priori* known tight support constraint unless otherwise indicated.

Figure 3 shows a block diagram of the conventional ITA. A starting guess for the object was created by assigning a constant amplitude and random phase distribution inside the object support. For the kth iteration, the object-domain function $g_k(\mathbf{x})$ is Fourier transformed, giving

$$G_k(\mathbf{u}) = |G_k(\mathbf{u})|\exp[i\theta_k(\mathbf{u})]. \tag{2}$$

This result is projected onto the measured magnitude constraint by imposing the measured magnitude while keeping the phase unchanged,

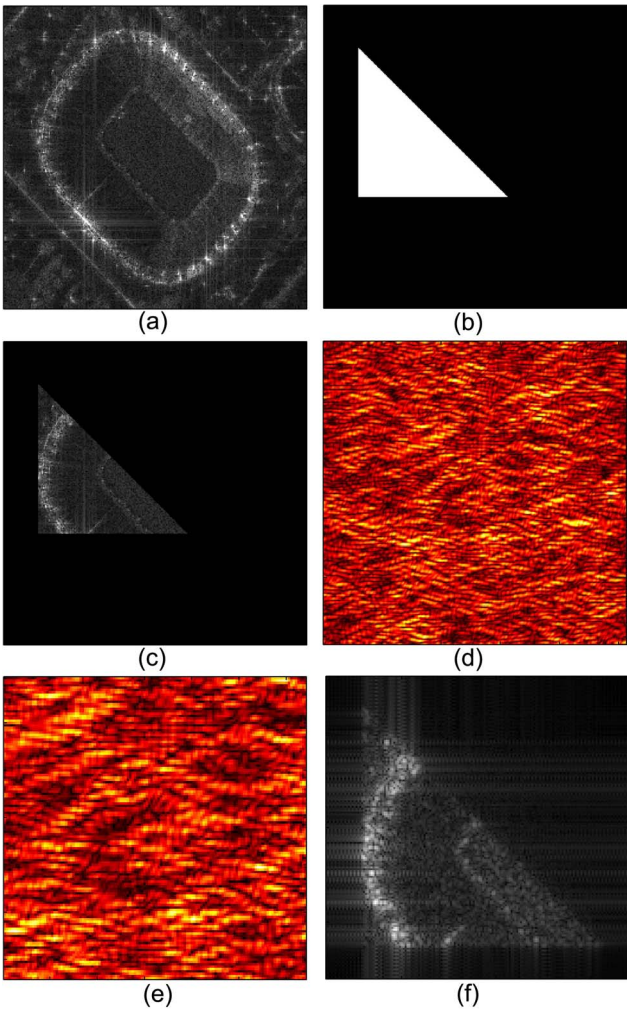

Fig. 2. (Color online) (a) Original SAR image. (b) Hard-edge object support. (c) 256×256 finite-support object and (d) its Fourier magnitude. Nonphysical wrap-around effects are present in (d) at the edge of the computational window. (e) 128×128 central portion of (d), which is used by the phase retrieval algorithms. (f) Diffraction-limited image magnitude; the 128×128 image was upsampled by $2\times$, and only a 160×160 pixel inset is displayed. Due to their large dynamic ranges, the square root of the magnitude is displayed in (a), (c), and (f).

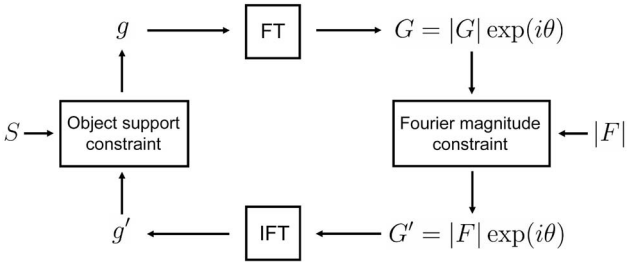

Fig. 3. Block diagram of the iterative transform algorithm.

$$G_k'(\mathbf{u}) = |F(\mathbf{u})|\exp[i\theta_k(\mathbf{u})], \tag{3}$$

where $\theta_k(\mathbf{u})$ is the far-field phase estimation for the kth iteration. Then an inverse Fourier transform is performed, giving $g_k'(\mathbf{x})$, and the object support constraint is used to form the new $g_{k+1}(\mathbf{x})$, as described later. The algorithm iterates over this cycle until an output that satisfies both sets of constraints is found or the algorithm

stagnates, making no further progress. Convergence of these algorithms is usually monitored with the object support error

$$E^2 = \frac{\sum_{\mathbf{x}\notin S} |g'(\mathbf{x})|^2}{\sum_{\mathbf{x}} |g'(\mathbf{x})|^2}, \tag{4}$$

where S is the set of points for which $g'(\mathbf{x})$ satisfies the object-domain constraints. For the example of Fig. 2(f), this metric is nonzero for the ideal diffraction-limited image ($E=0.203$) due to the presence of sidelobes outside the support.

A number of different techniques have been proposed to use the object constraints; among the most commonly used are the HIO [12] and error-reduction versions of the ITA.

The error-reduction algorithm imposes the support constraint by setting to zero any output values outside the support constraint (for the case of having only a support constraint),

$$g_{k+1}(\mathbf{x}) = \begin{cases} g'_k(\mathbf{x}), & \text{if } \mathbf{x} \in S \\ 0 & \text{if } \mathbf{x} \notin S \end{cases}, \tag{5}$$

and thus represents the projection onto the image constraints. It has been proven that this reduces E on every iteration [12], which unfortunately means that escaping local minima is impossible and makes the algorithm prone to stagnation.

Conversely, the HIO algorithm uses the previous input to drive the output to zero at points where the support constraint is violated,

$$g_{k+1}(\mathbf{x}) = \begin{cases} g'_k(\mathbf{x}), & \text{if } \mathbf{x} \in S \\ g_k(\mathbf{x}) - \beta g'_k(\mathbf{x}), & \text{if } \mathbf{x} \notin S \end{cases}, \tag{6}$$

where the feedback parameter β is typically assigned a value between 0.5 and 0.8. HIO has shown to be successful at escaping local minima and is a valuable tool for reconstructing complex-valued images.

The best results are usually obtained by using error reduction and HIO in combination, using HIO in most iterations to explore the solution space and periodically performing a few iterations of error reduction to help the current reconstruction settle down. Although this algorithm combination has proven to be robust, it still may get trapped in stagnation modes that are very persistent [13], especially when applied in its simplest form to complex-valued image reconstruction [10].

For this image-reconstruction problem we used an improved version of the ITA that includes an expanding tapering function on the Fourier magnitude [4,10]. In the first set of iterations, the measured Fourier magnitude is multiplied by a narrow tapering function, so that the phase retrieval algorithm is reconstructing a low-resolution image by only retrieving the phase over a small aperture. The tapering over that small aperture helps by reducing the sidelobes in the image. After a solution to this reduced problem is reached, which occurs many times faster than for the fully resolved reconstruction using the entire aperture, we slightly expand the tapering function and continue iterating. This cycle of expanding and iterating is continued until the entire computational window is included. With this modification, the ITA gradually bootstraps from a solution over a small aperture to successively larger ones, which substantially improves complex-valued image reconstruction [10].

We applied the expanding tapering ITA algorithm to reconstruct a fully resolved image from the simulated data shown in Fig. 2(e). Because the simulation of far-field data involves truncating the object FFT, the width of the impulse response is increased, and best results were obtained by using a support constraint that was dilated by one pixel. Twenty series of 45 HIO iterations with $\beta=0.7$ followed by 5 of error reduction were used, totaling 1000 ITA iterations. The tapering function used was the autocorrelation of a circle with diameter equal to a fraction of the total length of the computational window. For the first 50 iterations this fraction was equal to 0.3, and it was gradually increased by 0.1 after each series of iterations until a value of 1.5 was reached. Subsequent iterations were performed without any Fourier tapering function.

Figure 4(a) shows the reconstructed image obtained from the steps described above. The reconstruction resembles the ideal image shown in Fig. 2(f) but appears significantly noisy due to an imperfect phase estimate. The far-field phase error, $\Delta\phi(\mathbf{u}) = \theta_k(\mathbf{u}) - \phi(\mathbf{u})$, of the reconstructed far-field phase distribution, where $\phi(\mathbf{u})$ is the true phase, is shown in Fig. 4(b). Computation of the far-field phase error requires knowledge of the true phase and is thus available only for numerical simulations. It is computed by first registering the reconstruction $g(\mathbf{x})$ to the diffraction-limited image $f(\mathbf{x})$ to within a small fraction of a pixel [18] (in this case that fraction was 1/200), removing the global constant phase between them, and computing the difference of their far-field phases.

The fact that the phase error is worst at the edge of the computational window is partially explained by the fact that the correct far-field phase and amplitude do not wrap around, which is effectively seen by the FFT as an abrupt edge in Fourier space, spreading energy throughout image space. Since we are also imposing a hard-edge support constraint on object space, there is no solution that can satisfy the constraints in both domains. By continually reinforcing the hard-edge image support constraint, which imposes smoothness in the Fourier domain, the ITA smoothes out the phases at the edge of the computational window even though the measured magnitude is not continuous there. Of course this produces an error on the retrieved image because the phase of the true solution does not wrap around and is not smooth across the edge. This inconsistency significantly reduces the performance of the reconstruction algorithm. Note, however, that this is much less of a problem for a real nonnegative object whose Fourier transform is naturally smaller near the edge of the computational window.

There are a few techniques that could be used to avoid this wrapped-around phase problem: (1) The object support could be expanded so that it includes a greater fraction of the image sidelobes. Unfortunately, it has been shown that a tight support constraint is very important for a successful reconstruction of complex-valued images, and expanding the support even by a few pixels greatly

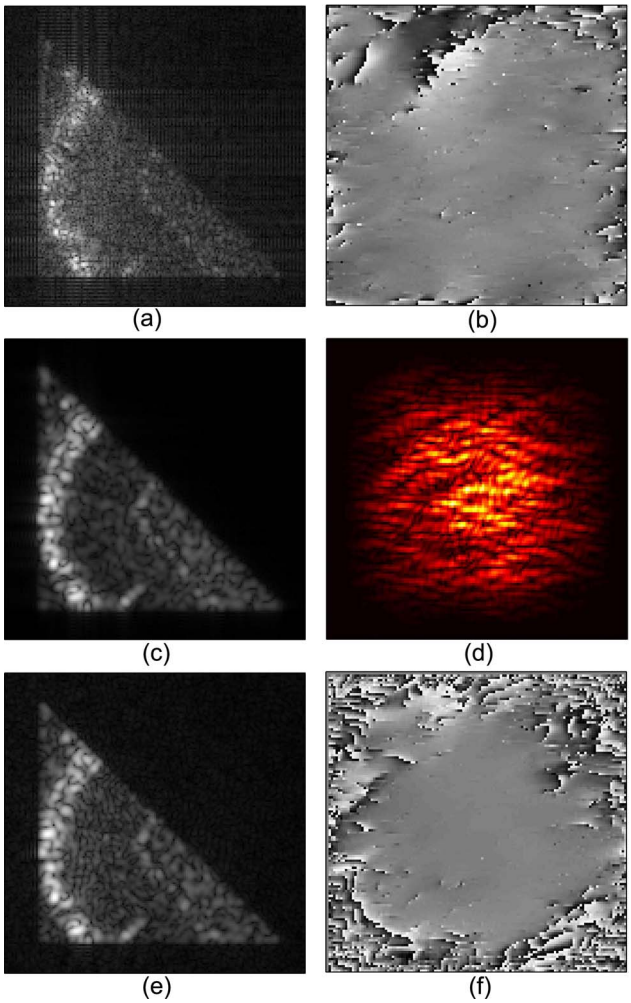

Fig. 4. (Color online) (a) Final reconstruction using the far-field magnitude shown in Fig. 2(e). (b) Far-field phase error $\Delta\phi(\mathbf{u})$ of (a). (c) Low-pass diffraction-limited image and (d) its Fourier magnitude. (e) Reconstruction from the low-pass far-field magnitude shown in (d). (f) Far-field phase error of (e). Phase is shown from $-\pi$ to π in (b) and (f).

hinders the algorithm performance [9]. Furthermore, since the sidelobes extend, at lower amplitudes, throughout all of image space, even an expanded support constraint will not include all the sidelobes. (2) A hard edge in Fourier space can be enforced by zero padding the measured amplitude. This approach was attempted with discouraging results. Without the opportunity to compensate the object hard-edge constraint by wrapping around the phases, the algorithm wandered around in the solution space, and the final reconstruction worsened significantly. (3) Perhaps the most effective technique is the application of a multiplicative weighted low-pass window in Fourier space [4,10]. If an appropriate weighting function is chosen, the image sidelobes can be greatly reduced without hindering the algorithm performance, but at the cost of a reconstruction having reduced resolution because the reconstructed image would be equivalent to that obtained with an apodized optical system. This may be understood as allowing the ITA to wrap around the phase in Fourier space while weighting down the areas at the edge of the computational window where most of the error will occur. It is emphasized that, since the phases at the edge are in error in this reconstruction, boosting up its high spatial frequencies by means of an inverse filter or by replacing the filtered magnitude with the original one does not fix the problem of the phase at the edges of the computational window.

An example using this third technique is shown in Figs. 4(c)–4(f). The Fourier transform shown in Fig. 2(e) was multiplied by a separable Hanning filter giving the weighted Fourier transform shown in Fig. 4(d). The corresponding diffraction-limited (low-pass-filtered) image, shown in Fig. 4(c), has greatly reduced sidelobes, but at the expense of poorer resolution as compared with the fully resolved diffraction-limited image [Fig. 2(f)]. Since the low-pass filtering further increases the size of the impulse response (besides the truncation of the object Fourier transform), the best reconstruction results were obtained by using a support constraint that was dilated by two pixels. The reconstruction after 1000 iterations and its far-field phase error are shown in Figs. 4(e) and 4(f), respectively. The phase error is small where the weighted Fourier magnitude is large and is large where the Fourier weighting is small, so the reconstructed image has moderately high quality.

3. IMPROVED PHASE RETRIEVAL WITH WEIGHTED PROJECTIONS

Since an important advantage of the potential application of CDI, both for x-ray and lensless laser imaging, is the acquisition of high-resolution images, it would be desirable to use all of the measured far-field intensity without sacrificing resolution with Fourier weighting. To circumvent the sidelobe problem, we propose a controlled extrapolation technique based on weighted projections in Fourier space. A weighted projection is achieved by substituting the Fourier projection step, given in Eq. (3) by [19]

$$G_k'(\mathbf{u}) = W(\mathbf{u})|F(\mathbf{u})|\exp[i\theta_k(\mathbf{u})] + [1 - W(\mathbf{u})]G_k(\mathbf{u}), \quad (7)$$

where $W(\mathbf{u})$ is the weighting function. Wherever $W(\mathbf{u}) = 1$, the Fourier magnitude constraint is enforced, while a value of $W(\mathbf{u}) = 0$ leaves both the magnitude and the phase of $G_k(\mathbf{u})$ unchanged, and intermediate values yield something in between. Equation (7) represents a relaxed projection onto the Fourier magnitude constraint [20] but where the relaxation parameter is a function of the spatial frequency coordinates $\mathbf{u}$. Setting the weighting function to zero has been used as an interpolation technique for areas where there is no confidence in the measured data due to bad or saturated pixels, low signal-to-noise ratio (SNR), or where no data were measured [8,19,21].

Chapman *et al.* successfully reconstructed a three-dimensional image from x-ray diffraction patterns taken at different sample orientation angles [8]. These diffraction patterns were processed to yield a Fourier amplitude data cube that had small gaps due to large rotation increments and a large 40° sector of missing data due to a limited range of sample orientations. In their reconstruction they used $W(\mathbf{u}) = 0$ to allow the algorithm to interpolate in the regions of missing data. Upon comparing reconstruction from several random starting guesses, they observed

that the recovered (interpolated) data were inconsistently reconstructed and averaged to zero.

Iterative extrapolation using a support constraint has been previously described to achieve super-resolution [22]. Gerchberg describes a technique in which a weight $W(\mathbf{u})=0$ is assigned outside of the measurement area during the iteration process, thus allowing the Fourier data to freely extend outside of the measurement window. However, super-resolution is known to be an ill-posed problem [23], and it will be more so if we attempt to simultaneously extrapolate to achieve super-resolution and retrieve the phase distribution.

In general the problem of data extrapolation is less robust than interpolation, and the computed extrapolation becomes less certain as we move farther from the measured data. Consequently we use a weighting function that does not allow the extrapolation to extend freely across the entire computational window. Use of a nonbinary weighting function outside of the measurement window allows the algorithm to slightly extrapolate the measured data. The purpose is not to achieve true super-resolution but to extrapolate the Fourier data just enough to make them consistent with the hard-edge object support constraint. This effectively expands the allowed solution space, so that the diffraction-limited solution is included without introducing a large number of free parameters.

For our extrapolation technique, the first step is to embed the measured data (which was previously truncated in the forward model to avoid the wrap-around edges) in a larger array of zeros. The weighting function $W(\mathbf{u})$, shown in Figs. 5(a) and 6(a), is set to one within the measurement area so that the measured magnitude is enforced. In the zero-padded region the weight starts at zero at the edge of the measured magnitude and gradually increases to unity at the edge of the computational window. This weighting function allows the algorithm to extend freely when close to the measured data and gradually imposes a zero amplitude when approaching the edge of the computational window, thus allowing an analytic continuation of the measured data while preventing wrap-around effects.

The extrapolation of the measured magnitude allows the far-field data to be consistent with the hard-edge object constraint and improves the reconstruction without sacrificing resolution or hindering the algorithm performance. The extent of the zero padding should be chosen considering that, while true sidelobes extend indefinitely, FFTs on a larger array increase both the computation time and the memory requirements.

The simulated 128×128 Fourier magnitude data array was embedded in a 256×256 array of zeros and used to reconstruct the image shown in Fig. 5(c) by using the ITA with weighted projections as described by Eq. (7). The weighting function shown in Fig. 5(a) was used when imposing Fourier constraints. A value of $W(\mathbf{u})=1$ was assigned within the 128×128 measurement area, and a separable raised-cosine half-period variation was assigned in the zero-padded area. A cut through the center of $W(\mathbf{u})$ is shown in Fig. 6(a). Series of HIO and error reduction iterations were implemented with an expanding Fourier taper function as previously described. Again, the best results were obtained by dilating the object support by one pixel. Notice that the reconstruction has significantly reduced sidelobes as compared with the diffraction-limited image in Fig. 2(f). Figure 5(b) shows the fourth root of the magnitude of the extrapolated far field. Notice that the extrapolated sidelobes protrude in directions perpendicular to the edges of the object support. Figure 5(d) shows the far-field phase error $\Delta\phi(\mathbf{u})$ of the extrapolated reconstruction with respect to the original 256×256 far

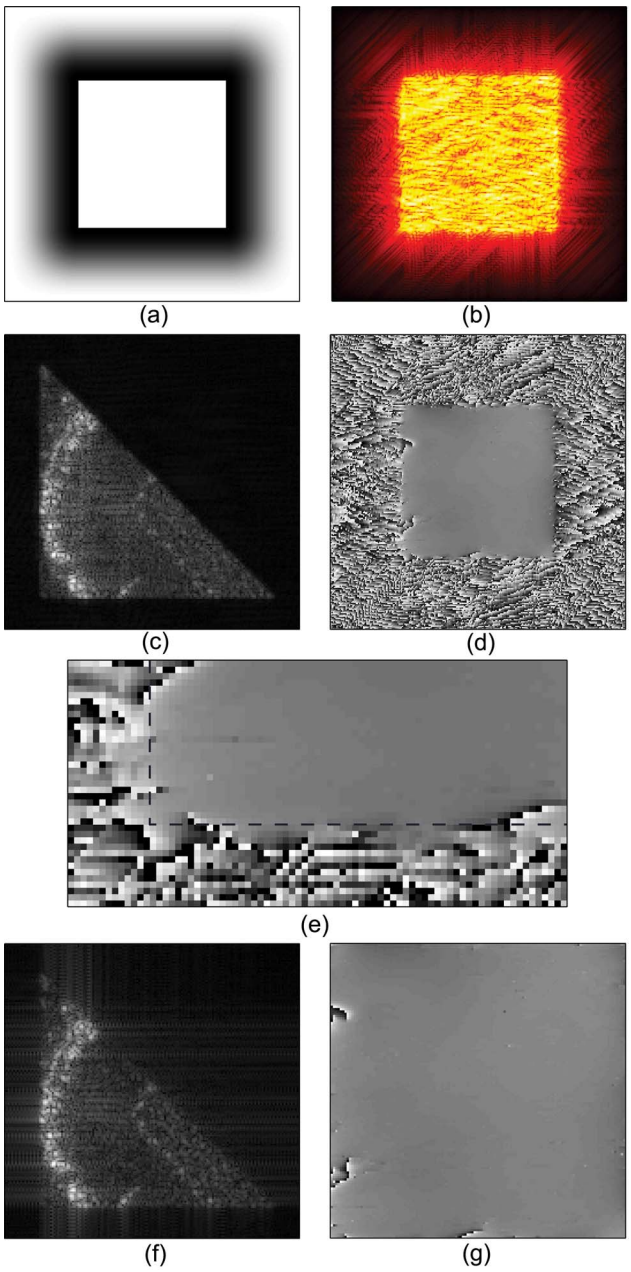

Fig. 5. (Color online) (a) 256×256 weighting function $W(\mathbf{u})$. (b) Fourth root of the far-field magnitude of the extrapolated reconstruction. (c) Reconstruction with the ITA with weighted projections with the far-field extrapolated data included. (d) 256×256 far-field phase error $\Delta\phi(\mathbf{u})$ of (c) with respect to object Fourier transform [Fig. 2(d)]. (e) 40×80 inset of (d); dashed line shows the edge of the measurement window. (f) Same as (c) but zeroing the far-field extrapolated data. (g) Far-field phase error within the original 128×128 measurement area. Phase is shown from $-\pi$ to π in (d), (e), and (g).

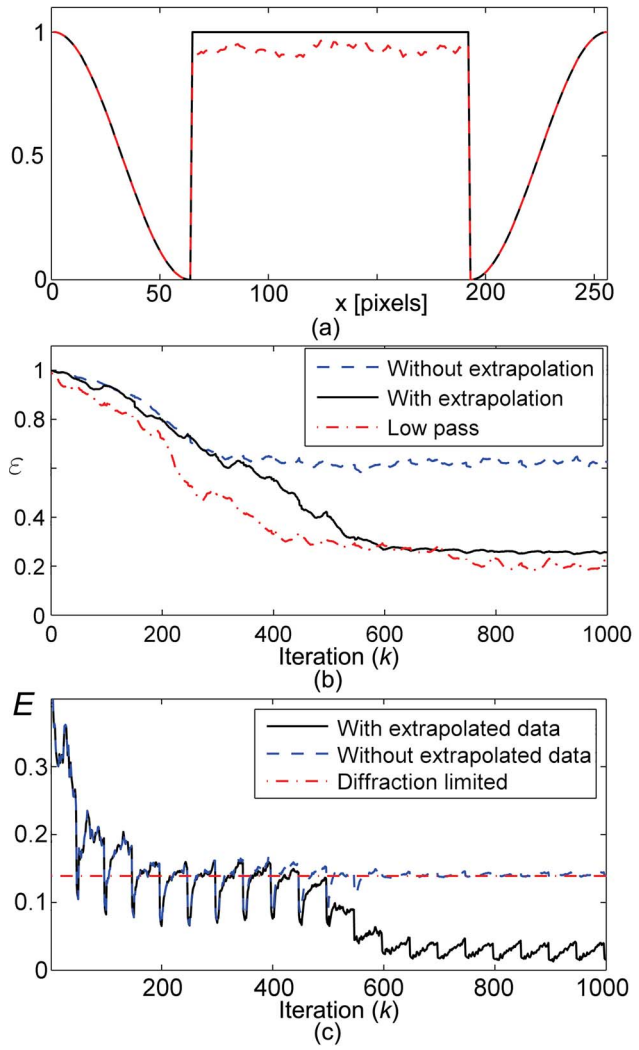

Fig. 6. (Color online) (a) Cut through the weighting function $W(\mathbf{u})$ (solid curve) and $W_\eta(\mathbf{u})$ for the weight modulation approach used for SNR=5 and η=0.1 (dashed curve). (b) NRMSE, ε, versus iteration number for the fully resolved (with and without use of extrapolation by weighted projections) and the low-pass reconstructions. (c) Object support error, E, versus iteration number for the reconstruction with the ITA with extrapolation by weighted projections. E was computed with and without the far-field extrapolated data for the same reconstruction. The support error, E=0.139, for the diffraction-limited image is shown for comparison purposes (horizontal dashed–dotted line).

field, the magnitude of which is shown in Fig. 2(d). A small subset of the far-field phase error is shown in Fig. 5(e). Notice that the retrieved phase is accurate for only a few pixels outside of the edge of the measurement window (indicated by a dashed line). This emphasizes the importance of weighting down the extrapolated amplitude on the iterative loop as we move farther from the measured data. This result also reveals that, although it was not the main purpose of this work, the analytic continuation achieved a small amount of true super-resolution.

For comparison with the diffraction-limited image, shown in Fig. 2(f), the reconstruction result shown in Fig. 5(f) was obtained by using the retrieved phases on the measurement area and zeroing out the extrapolated data. The improvement over the reconstruction shown in Fig. 4(a) is evident, and it has significantly higher resolution than the low-pass reconstruction shown in Fig. 4(e). Figure 5(g) shows the far-field phase error of the reconstruction within the 128×128 measurement window, which is significantly improved over those shown in Figs. 4(b) and 4(f). The retrieved far-field phase error is 0.043 waves root mean square (RMS) over the measurement window.

The normalized root mean square error (NRMSE) with respect to the diffraction-limited image was computed for every iteration as an objective means of comparison for the algorithms and to monitor the progress of the reconstruction. The NRMSE is defined by [24]

$$\varepsilon_k^2 = \min_{\alpha,\mathbf{x_0}} \left[\frac{\sum_{\mathbf{x}} |f(\mathbf{x}) - \alpha g_k(\mathbf{x}-\mathbf{x_0})|^2}{\sum_{\mathbf{x}} |f(\mathbf{x})|^2} \right], \tag{8}$$

where α and $\mathbf{x_0}$ are a multiplicative constant complex factor and a coordinate translation, respectively. The result of the minimization procedure described by Eq. (8) can be computed by finding the value and location of the peak value of the cross correlation of $f(\mathbf{x})$ and $g_k(\mathbf{x})$ to within a small fraction of a pixel [18,24].

Figure 6(b) shows a comparison of the NRMSE for the fully resolved, low-pass, and extrapolation reconstruction results versus the number of iterations. When computing the NRMSE of the extrapolation approach, any data outside of the far-field measurement area were ignored. Discarding the extrapolated data is necessary for comparison with the diffraction-limited image. Because the diffraction-limited image has zero energy outside of the measurement window, any extrapolated data, even if correct, will increase this error metric.

The NRMSE of the low-pass reconstruction, shown in Fig. 4(e), is computed with respect to the low-pass diffraction-limited image shown in Fig. 4(c). Notice that although the NRMSE of the low-pass reconstruction is lower than that obtained with the extrapolation, the former is computed with respect to a low-pass diffraction-limited image of decreased resolution.

Figure 6(c) shows the support error E, given by Eq. (4), versus iteration number for the extrapolated reconstruction. This error metric is commonly used to evaluate the progress of the reconstruction, and unlike the NRMSE with respect to the diffraction-limited image, it is available for reconstructions where the true phase is not known. For this reconstruction, the support error can be calculated using the current image estimate with or without the extrapolated data. Both of these calculations were performed and are displayed in Fig. 6(c).

During the first 45 HIO iterations of each series of 50 iterations, the support error tends to increase as the algorithm explores the solution space. This capability of escaping local minima is what makes the HIO a particularly valuable tool for phase retrieval. As the five error-reduction iterations are applied at the end of each series, the support error quickly drops and settles. The support error was the same with and without the extrapolated data for the first 200 iterations, on account of the nature of the expanding Fourier tapering approach. Before iteration 200, the Fourier magnitude was multiplied by the au-

tocorrelation of a circle with diameter equal to one half of the entire computational window; this tapering function goes to zero at the edge of the measurement window, so no extrapolation is performed up to this point. Upon further expansion of the tapering function, the algorithm gradually begins the extrapolation procedure, and the calculations with and without inclusion of the extrapolated data become increasingly distinct. The support error of the diffraction-limited image ($E=0.139$) is shown as well for comparison purposes. This error is lower than the one that was previously quoted ($E=0.203$) because the support used in the reconstruction was dilated by one pixel.

Without the far-field extrapolated data, as the tapering function was removed the support error tended to that of the diffraction-limited image. With the far-field extrapolated data, as the tapering function was removed the support error decreased, with values that are significantly lower than those of the diffraction-limited image. Notice in particular that the support error, E, and the NRMSE are strongly correlated. Thus the support error computed including the extrapolated data is in this case a valuable metric to evaluate the reconstruction performance.

To assess the effect of noise on the reconstructed image, we simulated photon noise in the measured intensity data (then took the square root to obtain the Fourier magnitude) and performed 1000 iterations of the ITA with extrapolation by weighted projections as described above. Figures 7(a)–7(c) show final reconstructions for average SNR of 100, 10, and 5, which correspond to an average of 10,000, 100, and 25 photons, respectively, per detector pixel. As expected, the algorithm performance degrades as the noise in the measurements increases, although the image is still recognizable with only 100 photons per pixel.

We can use the weighted projections approach to take advantage of the knowledge that the SNR is lower for pixels where the measured intensity is low, by applying a weight modulation within the measurement area that will aid the reconstruction for low photon counts.

Having a weighting function that has a slightly lower value where the SNR is low can help the reconstruction by allowing a slight modification of the far-field amplitude values. An alternative weighting function of the form

$$W_\eta(\mathbf{u}) = 1 - \eta + \eta \frac{|F(\mathbf{u})|}{\max_{\mathbf{u}}|F(\mathbf{u})|} \tag{9}$$

was used to improve the reconstruction for SNR=5, where $|F(\mathbf{u})|$ is the noisy measured data and $\eta<1$ is a constant.

A cut through this weighting function for noisy data, for the simulated data with SNR=5 and $\eta=0.1$, is shown in Fig. 6(a). Figure 7(d) shows the reconstruction result after 1000 iterations. Although significant improvement can be seen for SNR=5, the reconstruction does not improve when this is implemented for SNR of 10 or 100. Figure 7(e) shows the NRMSE of the reconstructions from noisy data as a function of the number of iterations.

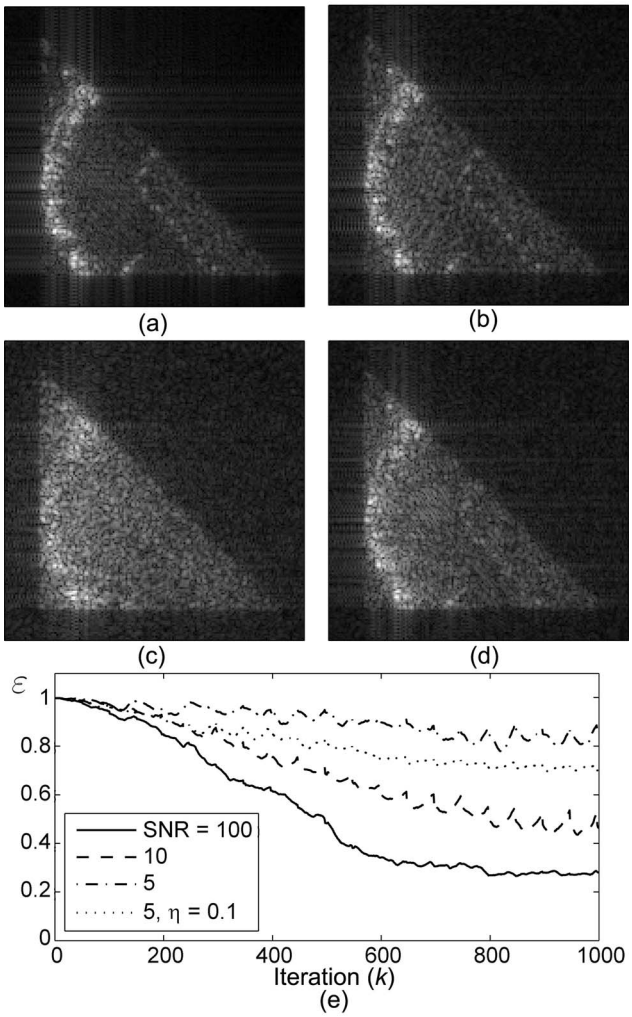

Fig. 7. Reconstruction with the ITA with weighted projections for an average SNR of (a) 100, (b) 10, and (c) 5. (d) Reconstruction for SNR=5 using the alternative weighting for noisy data ($\eta=0.1$). (e) NRMSE, ε, versus iteration number for reconstructions from noisy data.

4. CONCLUSION

An important problem arising in coherent computational image reconstruction is the presence of sidelobes that extend beyond the object support in the diffraction-limited image. These sidelobes are due to the finite extent of the detector array, and their inclusion in numerical simulation of data acquisition is crucial to assess the performance of reconstruction algorithms. By truncating the Fourier data computed from the ideal object, we make numerical simulations more realistic by including the effects of the sidelobes. It was found that this truncation significantly hinders the performance of phase retrieval algorithms because finite data in the Fourier domain is mathematically inconsistent with the continually reinforced hard-edged object-support constraint.

This problem has been reasonably well managed in the past by applying a suitable low-pass filter in the Fourier domain, which reduces the sidelobe amplitude but at the expense of image resolution. As an alternative, we developed a controlled analytic continuation (extrapolation) for use with ITAs by means of weighted projections in Fourier space. The technique allows for better consistency between the extrapolated Fourier data and the object support constraint and was found to significantly improve the

quality of the reconstruction without sacrificing resolution. We found this approach to be robust in the presence of noise.

Algorithms that perform analytic continuation by iterative methods have been proposed previously to achieve super-resolution [22]. Although super-resolution is not the goal of the proposed extrapolation, we found that the extrapolated data was accurate for just a few pixels from the measurement window. We would like to emphasize that in this work the Fourier data are only slightly extended to make the object support constraint work better.

Extrapolation by nonbinary weighted projections can be used whenever a significant portion of the data is missing. If there are large voids of data in the measured Fourier magnitude, as occurs in the case of three-dimensional x-ray imaging [8], then as one interpolates farther away from the measured data, the interpolation becomes less certain and errors are likely to increase. In these cases, use of a weighting function whose magnitude increases as we move farther from the data should prove to be useful.

Finally, we also used an alternative projection weight function with a smaller value where the SNR is low within the far-field measurement area to improve the reconstructions for low average SNR. This approach allows the ITA to slightly change the amplitude of the measured data and was found to increase the overall reconstruction performance for an average SNR of 5. Although this approach seems promising for photon-limited applications, further research is needed to determine the optimum weighting function, $W_\eta(\mathbf{u})$, and coefficient, η, depending on the SNR and noise statistics.

Portions of this paper were presented in [25].

REFERENCES

1. J. R. Fienup, "Reconstruction of a complex-valued object from the modulus of its Fourier transform using a support constraint," J. Opt. Soc. Am. A **4**, 118–123 (1987).
2. P. S. Idell, J. R. Fienup, and R. S. Goodman, "Image synthesis from nonimaged laser-speckle patterns," Opt. Lett. **12**, 858–860 (1987).
3. J. N. Cederquist, J. R. Fienup, J. C. Marron, and R. G. Paxman, "Phase retrieval from experimental far-field speckle data," Opt. Lett. **13**, 619–621 (1988).
4. J. R. Fienup, "Lensless coherent imaging by phase retrieval with an illumination pattern constraint," Opt. Express **14**, 498–508 (2006).
5. S. G. Podorov, K. M. Pavlov, and D. M. Paganin, "A non-iterative reconstruction method for direct and unambiguous coherent diffractive imaging," Opt. Express **15**, 9954–9962 (2007).
6. M. Guizar-Sicairos and J. R. Fienup, "Holography with extended reference by autocorrelation linear differential operation," Opt. Express **15**, 17592–17612 (2007).
7. E. N. Leith and J. Upatnieks, "Reconstructed wavefronts and communication theory," J. Opt. Soc. Am. **52**, 1123–1130 (1962).
8. H. N. Chapman, A. Barty, S. Marchesini, A. Noy, S. P. Hau-Riege, C. Cui, M. R. Howells, R. Rosen, H. He, J. C. H. Spence, U. Weierstall, T. Beetz, C. Jacobsen, and D. Shapiro, "High-resolution *ab initio* three-dimensional x-ray diffraction microscopy," J. Opt. Soc. Am. A **23**, 1179–1200 (2006).
9. R. G. Paxman, J. R. Fienup, and J. T. Clinthorne, "The effects of tapered illumination and Fourier intensity errors on phase retrieval," Proc. SPIE **828**, 184–189 (1987).
10. J. R. Fienup and A. M. Kowalczyk, "Phase retrieval for a complex-valued object by using a low-resolution image," J. Opt. Soc. Am. A **7**, 450–458 (1990).
11. J. R. Fienup, R. G. Paxman, M. F. Reiley, and B. J. Thelen, "3D imaging correlography and coherent image reconstruction," Proc. SPIE **3815**, 60–69 (1999).
12. J. R. Fienup, "Phase retrieval algorithms: a comparison," Appl. Opt. **21**, 2758–2769 (1982).
13. J. R. Fienup and C. C. Wackerman, "Phase-retrieval stagnation problems and solutions," J. Opt. Soc. Am. A **3**, 1897–1907 (1986).
14. M. Guizar-Sicairos and J. R. Fienup, "Complex valued object reconstruction from extrapolated intensity measurements," in *Frontiers in Optics*, OSA Technical Digest Series (Optical Society of America, 2006), paper FMI6.
15. From the ERIM (now General Dynamics, Ypsilanti Michigan) X-band DCS radar.
16. T. R. Crimmins, J. R. Fienup, and B. J. Thelen, "Improved bounds on object support from autocorrelation support and application to phase retrieval," J. Opt. Soc. Am. A **7**, 3–13 (1990).
17. S. Marchesini, H. He, H. N. Chapman, S. P. Hau-Riege, A. Noy, M. R. Howells, U. Weierstall, and J. C. H. Spence, "X-ray image reconstruction from a diffraction pattern alone," Phys. Rev. B **68**, 140101 (2003).
18. M. Guizar-Sicairos, S. T. Thurman, and J. R. Fienup, "Efficient subpixel image registration algorithms," Opt. Lett. **33**, 156–158 (2008).
19. J. R. Fienup, "Phase-retrieval algorithms for a complicated optical system," Appl. Opt. **32**, 1737–1746 (1993).
20. A. Levi and H. Stark, "Restoration from phase and magnitude by generalized projections," in *Image Recovery: Theory and Application*, H. Stark, ed. (Academic, 1987), pp. 277–320.
21. J. R. Fienup and J. D. Gorman, "Image reconstruction for an aberrated amplitude interferometer with a partially-filled aperture," in *NOAO-ESO Conference on High-Resolution Imaging by Interferometry: Ground-Based Interferometry at Visible and Infrared Wavelengths*, F. Merkle, ed. (1988), pp. 293–301.
22. R. W. Gerchberg, "Super-resolution through error energy reduction," Opt. Acta **21**, 709–720 (1974).
23. L. S. Joyce and W. L. Root, "Precision bounds in superresolution processing," J. Opt. Soc. Am. A **1**, 149–168 (1984).
24. J. R. Fienup, "Invariant error metrics for image reconstruction," Appl. Opt. **36**, 8352–8357 (1997).
25. M. Guizar-Sicairos and J. R. Fienup, "Iterative extrapolation for image reconstruction by phase retrieval from truncated far field intensity measurements," in *Coherence 2007: International Workshop on Phase Retrieval and Coherent Scattering* (ALS Communications, 2007), poster P10.