Quantitative phase reconstruction from input-output intensities using a shift-invariant linear imaging system

Rotha P. Yu* and David M. Paganin
School of Physics, Monash University, Victoria 3800, Australia
(Received 22 December 2008; published 1 May 2009)

We formulate an iteration scheme for quantitative two-dimensional phase retrieval of complex scalar wave fields from input-output intensity profiles using a coherent rotationally symmetric shift-invariant linear optical imaging system. This method is a generalization of our previous method for quantitative single-image pure phase reconstruction using an imperfect shift-invariant linear imaging system. The method is somewhat analogous to the Gerchberg-Saxton iterative phase-retrieval algorithm, but typically converges much more rapidly. We have demonstrated the efficacy of the method using simulated data in the presence of noise.

DOI: 10.1103/PhysRevA.79.053803 PACS number(s): 42.30.Rx, 02.30.Zz, 42.30.Wb

I. INTRODUCTION

The problem of coherent quantitative optical phase imaging seeks to solve for a two-dimensional monochromatic complex scalar wave field given one or more intensity maps over a given detector plane. Here the wave field is considered to be complex (with amplitude and phase), with the detected quantity being the intensity (squared modulus) of the wave field over the surface of the detector.

This raises a fundamental "phase-retrieval" question concerning determination of the phase of the complex wave field, from noninterferometric intensity measurements. Since phase is an intrinsic property of any complex wave field, and at optical and higher frequencies direct physical measurement of the phase is extremely difficult, the phase problem occurs in many well-known systems such as x-ray crystallography (see e.g., [1]), electron crystallography [2], and coherent diffractive imaging (see e.g., [2–7]).

The phase problem arises as a result of trying to recover "lost" phase information, due to physical measurements that register only the intensity of a given field (see e.g., [8,9]). In general it is impossible to obtain both the amplitude and phase information of the wave field given only a single intensity map. To remedy this, one may make *a priori* assumptions about the wave field, and/or take additional optical data.

As an example of this course of action, let us restrict consideration to the coherent-imaging scenario sketched in Fig. 1. Here, a coherent-imaging system converts the input complex field Ψ_{in} into the output complex field Ψ_{out}. Assume that one has knowledge of both the input intensity $I_{\text{in}}=|\Psi_{\text{in}}|^2$ and the output intensity $I_{\text{out}}=|\Psi_{\text{out}}|^2$. In this case one has the problem of phase reconstruction [i.e., determining $\arg(\Psi_{\text{in}})$ and $\arg(\Psi_{\text{out}})$] given both the input and output intensity profiles. A well-known method for attacking this problem, for the special case where the "imaging" system corresponds to far-field diffraction, is the iterative projection algorithm of Gerchberg and Saxton (GS) [10].

In its original form, the GS algorithm involves Fourier transformations between real and reciprocal spaces, together with modulus projections over the input and output planes [10]. At each step of this iteration scheme, one replaces the modulus of the current iterate of the retrieved complex wave field with the measured modulus. Since its original formulation in 1972, many improvements have been made to this phase-retrieval scheme. For example a more useful GS-type algorithm has been implemented to solve the phase problem, where only a subset of the object is known (specifically a region within which the object is known to be entirely contained) [5,11,12]. Fienup's identification of the key role of such "support" information [11], in the context of phase retrieval, has formed a key ingredient in the continued successful development of iterative approaches to the phase problem. For further information regarding the numerous advances flowing from the seminal phase-retrieval work of Gerchberg and Saxton, see the excellent review article by Spence [2], together with references therein.

Such projection methods may be inefficient due to slow convergence, often requiring of the order of thousands of iterations in order to achieve a satisfactory reconstruction. Against this, one may consider noniterative and direct methods of phase retrieval, which perform the reconstruction in a "single shot." For example, one has the noniterative approaches to the phase problem published by Nakajima [13,14], Podorov *et al.* [15], Quiney *et al.* [16], Guizar-Sicairos and Fienup [17,18], and Martin and Allen [19]. Fourier holography should also be mentioned in this regard [20–22].

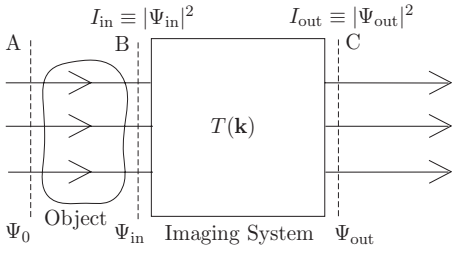

FIG. 1. Three sequential stages of a shift-invariant coherent optical imaging system—the complex input wave field Ψ_{in}, the imperfect imaging system characterized by coherent transfer function $T(\mathbf{k})$ [see Eq. (5)], and the output wave field Ψ_{out}. The phase-retrieval problem, considered in the present paper, is to determine Ψ_{out} given $|\Psi_{\text{in}}|$, $|\Psi_{\text{out}}|$, and T.

*rotha.yu@sci.monash.edu.au

In digital Fourier holography, one obtains the output far-field intensity pattern of the combined object wave field and the reference wave field at the image plane. Given the output intensity pattern (or hologram)

$$I_{\text{out}} = |\Psi_{\text{ref}}|^2 + |\Psi_{\text{out}}|^2 + \Psi_{\text{ref}}^* \Psi_{\text{out}} + \Psi_{\text{ref}} \Psi_{\text{out}}^*, \tag{1}$$

one obtains the output wave field as

$$\Psi_{\text{out}} = [I_{\text{out}} \Psi_{\text{ref}}]_{\text{filtered}} / |\Psi_{\text{ref}}|^2, \tag{2}$$

where Ψ_{ref} is a known reference wave field (any known wave field, but most commonly either a plane or a spherical wave), and for $[I_{\text{out}}\Psi_{\text{ref}}]_{\text{filtered}}$ the first, second, and the fourth terms on the right of Eq. (1) have been filtered out using the Fourier representation of the diffracted intensity (for a detailed review on digital holographic methods, see e.g., [23]).

Other important methods in noniterative quantitative phase imaging (for, in particular, light optics) are Fourier phase microscopy (FPM), Hilbert phase microscopy (HPM), and diffraction phase microscopy (DPM); for an outline, see e.g., Popescu [24]. Experimental work by Popescu indicates that these methods are increasingly useful for imaging of live biological samples, such as cell visualization and morphometric applications. In FPM, four different holograms are obtained using the unscattered beam as the reference wave field. To obtain the four different "phase stepped" holograms, the unscattered wave field is phase shifted by 0, $\pi/2$, π, and $3\pi/2$, respectively, using a programmable phase modulator. The phase $\Phi_{\text{out}} = \arg(\Psi_{\text{out}})$ of the scattered beam is calculated from the four holograms I_0, $I_{\pi/2}$, I_π, and $I_{3\pi/2}$ (the subscript indicates the amount of phase shift), as

$$\Phi_{\text{out}} = \tan^{-1}\left(\frac{I_{3\pi/2} - I_{\pi/2}}{I_0 - I_\pi}\right), \tag{3}$$

where the phase and the holograms are functions of the x and y coordinate positions. In HPM, the reference wave field is identical to the illumination wave field (this can be effected, for example, using a one→two-fiber splitter). The phase of the output wave field Φ_{out} is obtained from I_{out} in Eq. (1) using a high pass filter and Hilbert transform. Finally DPM is similar to FPM in the way that the unscattered beam is used as the reference wave field; however, a phase grating combined with a spatial filtering system is used to separate the scattered and the unscattered beams. The result is a HPM with common path geometry similar to that of FPM. For detailed descriptions of these three methods and their applications to cell biology, we again refer the reader to the work by Popescu [24].

For another example of a noniterative approach to the phase problem, a rapid quantitative method can be achieved by the use of noninterferometric approaches which seek a direct solution to the equation of evolution of the intensity of the complex wave field, such as solution to the transport-of-intensity equation in the context of quantitative phase-contrast imaging [25–27].

A general characteristic of phase-contrast imaging is the use of imperfections, in a coherent-imaging system, to render the phase of the input field visible as intensity variations in the corresponding output field (cf. [28]). Such "imperfections" [i.e., a transfer function T that is not equal to unity; cf. Eq. (5) below] are needed because a perfect imaging system, here defined as one that reproduces the input intensity over the output plane (up to multiplicative and transverse scale factors, and to a given resolution), is insensitive to the phase of the input field (cf. Fig. 1).

In a previous work [29], we have formulated an iteration scheme for quantitative single-image phase reconstruction of pure phase objects (i.e., objects over whose exit plane B one has uniform unit amplitude, cf. Fig. 1) using the output intensity I_{out} yielded by a coherent rotationally symmetric shift-invariant linear imaging system. The method has been successfully applied to simulated data of a general pure phase object, without recourse to the commonly assumed weak phase approximation [30]. In this paper we show that our iteration scheme for the pure phase object can be generalized to the case of objects where the coherent input amplitude $|\Psi_{\text{in}}|$ is not uniform in modulus, over the nominal exit plane B of the object (this coincides with the entrance plane of the coherent-imaging system; cf. Fig. 1).

The contents of the paper are summarized as follows. Section II briefly reviews the theory of imperfect shift-invariant coherent linear optical imaging systems, while Sec. III formulates our generalized theory of phase retrieval and subsequently applies it to simulated numerical examples. Finally, discussions and conclusions are presented in Sec. IV.

II. SHIFT-INVARIANT LINEAR IMAGING SYSTEMS

In this paper we adopt the following simplified model of coherent optical imaging. With reference to Fig. 1, one has the forward-propagating monochromatic complex disturbance Ψ_0, which impinges upon the nominal planar entrance surface A of an elastically scattering object. The resulting monochromatic wave field at the exit plane B of the object constitutes the wave field Ψ_{in} that is input into a coherent-imaging system. This system produces the wave field Ψ_{out} as output, over the plane C. Harmonic time dependence is suppressed throughout.

We restrict ourselves to coherent optical imaging systems that are both linear and shift-invariant. The imaging scenario is "coherent" in the sense that the object is illuminated by a monochromatic wave field, with the corresponding output being monochromatic and of the same frequency as the input field. A system is linear if the complex output wave field, Ψ_{out}, can be written as the following integral transform involving Ψ_{in} and the complex impulse response h of the system (Green's function, propagator, or complex point-spread function)

$$\Psi_{\text{out}}(\mathbf{x}) = \int \Psi_{\text{in}}(\mathbf{x}') h(\mathbf{x};\mathbf{x}') d\mathbf{x}'. \tag{4}$$

Here $\mathbf{x} \equiv (x,y)$ and $\mathbf{x}' \equiv (x',y')$ denote two-dimensional real-space coordinates, and $h(\mathbf{x};\mathbf{x}')$ is the complex impulse response of the system at $\mathbf{x}$ from the input at $\mathbf{x}'$ [31]. The system is shift-invariant if the complex response, h, of the system depends only on relative coordinates, i.e., $h(\mathbf{x};\mathbf{x}') = h(\mathbf{x}-\mathbf{x}')$. In such a case, Eq. (4) is the convolution of the input wave field with the complex impulse response. The

convolution theorem of Fourier analysis then allows one to write down the following Fourier representation relating the output wave field $\Psi_{\text{out}}(\mathbf{x})$ to the input wave field $\Psi_{\text{in}}(\mathbf{x})$ as (see e.g., [31,32])

$$\mathcal{F}\Psi_{\text{out}}(\mathbf{x}) = T(\mathbf{k})\mathcal{F}\Psi_{\text{in}}(\mathbf{x}), \tag{5}$$

where $\mathbf{k} \equiv (k_x, k_y)$ represents the Fourier-space coordinates corresponding to $\mathbf{x}$, $T(\mathbf{k})$ is the coherent transfer function (which is the Fourier transform of the corresponding complex point-spread function or Green's function), and $\mathcal{F}$ denotes Fourier transform with respect to x and y.

With the corresponding inverse Fourier transform operator $\mathcal{F}^{-1}$, we obtain the following expression for the output wave field:

$$\Psi_{\text{out}}(\mathbf{x}) = \mathcal{F}^{-1}T(\mathbf{k})\mathcal{F}\Psi_{\text{in}}(\mathbf{x}). \tag{6}$$

Note that in the above expression, all operators are considered to act from right to left. For a sufficiently well-behaved shift-invariant linear imaging system in which the transfer function has no zeros, we may write (see e.g., [32])

$$T(\mathbf{k}) = \exp\left(i\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\tilde{\alpha}_{mn}k_x^m k_y^n\right), \tag{7}$$

where the set of complex numbers $\{\tilde{\alpha}_{mn}\}$ are the coefficients characterizing the imperfections of the shift-invariant coherent-imaging system. We re-iterate that such "imperfections" are necessary in order for a linear shift-invariant system to be sensitive to the phase of $\Psi_{\text{in}}(\mathbf{x})$, which in turn constitutes a necessary (but not sufficient) condition for solution of the associated phase-retrieval problem of determining the phase of $\Psi_{\text{in}}(\mathbf{x})$ from noninterferometric measurements of wave field moduli (i.e., given $I_{\text{in}} \equiv |\Psi_{\text{in}}|^2$ and $I_{\text{out}} \equiv |\Psi_{\text{out}}|^2$; cf. Fig. 1).

III. PHASE-RETRIEVAL FORMALISM

A. General formulation

In formulating our phase-retrieval scheme, we follow [32] by writing the coherent transfer function as a Taylor series,

$$T(\mathbf{k}) \equiv 1 + i\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\alpha_{mn}k_x^m k_y^n. \tag{8}$$

This serves to define the set of complex numbers $\{\alpha_{mn}\}$ as the coefficients characterizing the imperfections of the imaging system. By series expansion of Eq. (7) and equating with Eq. (8) for terms with the same power of k_x and k_y, one may readily relate the two sets of aberration coefficients $\{\tilde{\alpha}_{mn}\}$ and $\{\alpha_{mn}\}$ [32].

For the transfer function Eq. (8) applied to a rotationally symmetric linear shift-invariant coherent-imaging system via Eq. (6), one can show (see our previous paper [29]) that the output intensity, $I_{\text{out}} = |\Psi_{\text{out}}|^2$, is related to the input intensity, $I_{\text{in}} = |\Psi_{\text{in}}|^2$, as

$$I_{\text{out}}^{1/2}e^{i\Delta\Phi} = I_{\text{in}}^{1/2} + \sum_{n=1}^{\infty}\frac{i\alpha_{2n}(A_{2n} + iB_{2n} + iC_{2n})}{(2\pi i)^{2n}}, \tag{9}$$

where $\Delta\Phi \equiv \arg(\Psi_{\text{out}}) - \arg(\Psi_{\text{in}})$, and A_{2n}, B_{2n}, together with C_{2n}, are real-valued functions recursively defined in the following manner [29]. Let

$$A_2 \equiv \nabla^2 A_0 - A_0 \nabla\Phi \cdot \nabla\Phi, \tag{10}$$

$$B_2 \equiv 2\nabla A_0 \cdot \nabla\Phi, \tag{11}$$

$$C_2 \equiv A_0\nabla^2\Phi, \tag{12}$$

where $A_0 \equiv \sqrt{I_{\text{in}}}$ and $\Phi = \arg(\Psi_{\text{in}})$ is the phase of the input wave field. For successive values of the subscript n, we have

$$A_n \equiv \nabla^2 A_{n-2} - A_{n-2}\nabla\Phi \cdot \nabla\Phi - B_{n-2}\nabla^2\Phi - 2\nabla B_{n-2} \cdot \nabla\Phi - C_{n-2}\nabla^2\Phi - 2\nabla C_{n-2} \cdot \nabla\Phi, \tag{13}$$

$$B_n \equiv \nabla^2 B_{n-2} - B_{n-2}\nabla\Phi \cdot \nabla\Phi + A_{n-2}\nabla^2\Phi - C_{n-2}\nabla\Phi \cdot \nabla\Phi + 2\nabla A_{n-2} \cdot \nabla\Phi, \tag{14}$$

$$C_n \equiv \nabla^2 C_{n-2}. \tag{15}$$

Note that n in Eqs. (13)–(15) takes the values $4, 6, 8, \ldots, \infty$.

For a rotationally symmetric linear shift-invariant coherent-imaging system, Eq. (9) is still an exact (albeit highly nonlinear) equation that relates the output intensity, $I_{\text{out}} = |\Psi_{\text{out}}|^2$, to both the input intensity, $I_{\text{in}} = |\Psi_{\text{in}}|^2$, and the input phase, $\Phi = \arg(\Psi_{\text{in}})$. Separating real and imaginary parts, one may write Eq. (9) as

$$I_{\text{out}}^{1/2}\cos(\Delta\Phi) = I_{\text{in}}^{1/2} - \sum_{n=1}^{\infty}\frac{\alpha_{2n}^{\text{R}}(B_{2n} + C_{2n}) + \alpha_{2n}^{\text{I}}A_{2n}}{(2\pi i)^{2n}}, \tag{16}$$

$$I_{\text{out}}^{1/2}\sin(\Delta\Phi) = \sum_{n=1}^{\infty}\frac{\alpha_{2n}^{\text{R}}A_{2n} - \alpha_{2n}^{\text{I}}(B_{2n} + C_{2n})}{(2\pi i)^{2n}}, \tag{17}$$

where the superscripts R and I stand for real and imaginary parts, respectively. Equation (16) can be rewritten as

$$\sum_{n=1}^{\infty}\frac{\alpha_{2n}^{\text{R}}}{(2\pi i)^{2n}}C_{2n} = I_{\text{in}}^{1/2} - I_{\text{out}}^{1/2}\cos(\Delta\Phi) - \sum_{n=1}^{\infty}\frac{\alpha_{2n}^{\text{R}}B_{2n} + \alpha_{2n}^{\text{I}}A_{2n}}{(2\pi i)^{2n}}, \tag{18}$$

where the left side contains the "linear" part, and is given in terms of the input intensity, the output intensity, and the residual nonlinear summation on the right side.

In our previous work [29], we solved Eq. (18) for the case of a pure phase object, given a single phase-contrast image [i.e., $I_{\text{out}}(x,y)$ is given]. In such a case $A_0 = \sqrt{I_{\text{in}}} = 1$. Using Eqs. (12) and (15), we then have

$$C_{2n} = \nabla^{2n}\Phi. \tag{19}$$

Therefore for the case of a pure phase object, the phase is related directly to the real-valued functions C_{2n}. For a general complex scalar wave field, as considered in the present

paper, $A_0=\sqrt{I_{\text{in}}}$ is not uniform. In what follows we describe a method of solution applicable to such a case.

B. Complex wave field phase reconstruction

Here we solve Eq. (18) where both I_{in} and I_{out} are known single-valued and differentiable but otherwise arbitrary nonzero real-valued functions. This is done by first noting that using Eq. (15), we have $C_{2n}=\nabla^2 C_{2n-2}=\nabla^4 C_{2n-4}=\cdots=\nabla^{2n-2}C_2$. Using the Fourier derivative theorem, we have

$$C_{2n}=\nabla^{2n-2}C_2=\mathcal{F}^{-1}(-2\pi i\mathbf{k}\cdot\mathbf{k})^{n-1}\mathcal{F}C_2. \tag{20}$$

By substituting Eq. (20) into Eq. (18), we find that at the kth iteration

$$(C_2)_k=-\mathcal{F}^{-1}\left[\frac{\mathcal{F}[L_{k-1}+I_{\text{in}}^{1/2}]}{g(\mathbf{k})}\right], \tag{21}$$

where

$$g(\mathbf{k})\equiv\sum_{n=1}^{\infty}\frac{\alpha_{2n}^{\text{R}}(\mathbf{k}\cdot\mathbf{k})^{n-1}}{4\pi^2}, \tag{22}$$

$$L\equiv -I_{\text{out}}^{1/2}\cos(\Delta\Phi)-\sum_{n=1}^{\infty}\frac{\alpha_{2n}^{\text{R}}B_{2n}+\alpha_{2n}^{\text{I}}A_{2n}}{(2\pi i)^{2n}}. \tag{23}$$

Once C_2 at the kth iteration is known, we solve for the input phase using Eq. (12) (specifically $C_2=\sqrt{I_{\text{in}}}\nabla^2\Phi_{\text{in}}$). The input phase at the kth iteration is obtained via

$$\Phi_k=\mathcal{F}^{-1}\left\{\frac{[-4\pi^2\mathbf{k}\cdot\mathbf{k}-i\alpha]\mathcal{F}\left[\frac{C_2}{\sqrt{I_{\text{in}}}}+i\alpha\Phi_{k-1}\right]}{16\pi^4(\mathbf{k}\cdot\mathbf{k})^2+\alpha^2}\right\}, \tag{24}$$

where we have introduced regularization in Eq. (24), by adding $i\alpha\Phi\sqrt{I_{\text{in}}}$ (α is a small positive real parameter) to both sides of Eq. (12), to cope with the divergence due to the denominator at the Fourier-space origin. Note that the regularization is only made to Eq. (24), but not to Eq. (21) since the denominator of Eq. (21) does not vanish at the Fourier-space origin [since we have taken $(\mathbf{k}\cdot\mathbf{k})^0=1$ at $\mathbf{k}\cdot\mathbf{k}=0$]. Note further that Eq. (24) reduces to Eq. (15) in Ref. [29] for the case of a pure phase object upon setting $I_{\text{in}}=1$ (except for the regularization parameter α).

The stability and convergence of the algorithm for the case of an arbitrary complex object discussed here is similar to the case of a pure phase object discussed earlier [29]. Our argument is as follows. We solve Eqs. (21) and (24) in an iteration scheme where we assume that the nonlinear term is negligible at the first iteration. In successive iterations, the nonlinear terms are obtained using the input phase retrieved from the previous iteration. In such a case we are not actually solving the equation in the presence of the nonlinear term, but the nonlinear term is actually given. Therefore the stability of the solutions to Eqs. (21) and (24) can be understood by considering only the linear terms. If $g(\mathbf{k})$ is nonvanishing for all nonzero spatial frequencies, the solution to Eqs. (21) and (24) using our iteration scheme is unique up to a meaningless additive constant in the input phase. On the other hand if there are nonzero spatial frequencies at which $g(\mathbf{k})=0$, then one needs to take additional images and compute the outer bracket within the inverse Fourier transform in Eq. (21) using a suitable weighted average (see, e.g., [32,33]). As a specific example of such a strategy, given two different estimated solutions to Eq. (21), one can form C_2 from the average

$$(C_2)^{(\text{ave})}=-\mathcal{F}^{-1}[D^{(\text{ave})}], \tag{25}$$

where $D^{(\text{ave})}$ is obtained from two data sets as

$$D^{(1)}=\frac{\mathcal{F}[\mathcal{H}^{(1)}]}{g^{(1)}}\quad\text{and}\quad D^{(2)}=\frac{\mathcal{F}[\mathcal{H}^{(2)}]}{g^{(2)}},$$

with $\mathcal{H}\equiv L_{k-1}+I_{\text{in}}^{1/2}$. We may write $D^{(\text{ave})}$ using the following Fourier-space weighting formulas:

$$\begin{aligned}D^{(\text{ave})}&=\beta D^{(1)}+(1-\beta)D^{(2)},\\&=\frac{|g^{(1)}|^2D^{(1)}}{|g^{(1)}|^2+|g^{(2)}|^2}+\frac{|g^{(2)}|^2D^{(2)}}{|g^{(1)}|^2+|g^{(2)}|^2},\\&=\frac{(g^{(1)})^*g^{(1)}D^{(1)}}{|g^{(1)}|^2+|g^{(2)}|^2}+\frac{(g^{(2)})^*g^{(2)}D^{(2)}}{|g^{(1)}|^2+|g^{(2)}|^2},\\&=\frac{(g^{(1)})^*\mathcal{F}[\mathcal{H}^{(1)}]}{|g^{(1)}|^2+|g^{(2)}|^2}+\frac{(g^{(2)})^*\mathcal{F}[\mathcal{H}^{(2)}]}{|g^{(1)}|^2+|g^{(2)}|^2}.\end{aligned} \tag{26}$$

Notice that the real coefficients $\beta\equiv|g^{(1)}|^2/[|g^{(1)}|^2+|g^{(2)}|^2]$ and $1-\beta$ in the above equations sum to unity. Substituting Eq. (26) into Eq. (25) we obtain the equation for the average C_2 as

$$(C_2)^{(\text{ave})}=-\mathcal{F}^{-1}\left[\frac{(g^{(1)})^*\mathcal{F}[\mathcal{H}^{(1)}]+(g^{(2)})^*\mathcal{F}[\mathcal{H}^{(2)}]}{|g^{(1)}|^2+|g^{(2)}|^2}\right]. \tag{27}$$

For the above equation, note that if the value of $g^{(1)}$ vanishes at a particular Fourier-space coordinate (k_x,k_y), the solution of C_2 is obtained from the second value $g^{(2)}$ alone, and vice versa. Note also, in the above context of Fourier-space "splicing," the related statistically principled analysis of Huang and Anastasio [34].

The convergence of our iteration scheme may be studied by monitoring the change in physical quantities (such as the expectation value of the retrieved input phase gradient) or the errors of the magnitude of the retrieved output wave field to see if the errors are progressively smaller at each successive iteration. If this is the case, then in the sense of the Cauchy convergence criterion, the iteration converges.

C. Numerical example of phase reconstruction

For phase reconstruction of a general complex wave field, we consider simulated data taken from our multislice simulations of transmission electron microscopy (for references on the multislice method, see, e.g., [30,35,36]). While this simulated example is drawn from the field of transmission electron microscopy, we note that the method developed here is more generally applicable, for example to coherent complex scalar x-ray, visible-light, and neutron wave fields. Our multislice simulates a monochromatic planar electron beam

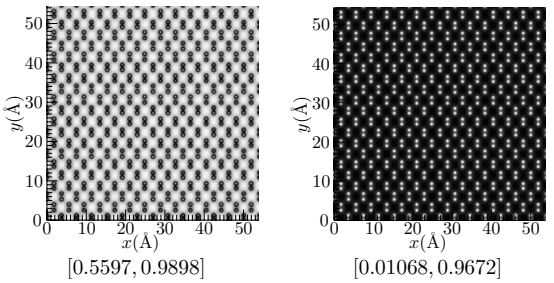

FIG. 2. The exit surface wave field of a 110 silicon crystal from our multislice simulation. The intensity $I_{\text{in}}(\mathbf{x})$ is shown on the left, with the phase $\Phi_{\text{in}}(\mathbf{x})$ on the right. The intensity (left) is in arbitrary units, whereas the phase (right) is in units of radians. Numbers below each image indicate the range of the gray scale value of each image. Dimensions of both images are 53.75 Å × 54.3 Å.

of wavelength $\lambda = 0.02508$ Å (electron energy 200 KeV) illuminating a 53.75 Å × 54.3 Å silicon crystal (3.8396 Å × 5.43 Å per unit cell) with the optical axis of the electron beam in the 110 silicon crystal direction. Numerical simulations were performed using 256 × 256 pixel spatial grids. After passing through ten layers of the multislice, totaling 19.198 Å thick (1.9198 Å thick per slice), the simulated electron wave field at the exit surface of the silicon crystal exhibits intensity and phase variations as shown in Fig. 2, with the intensity shown on the left and phase on the right. This exit wave field is the input wave field Ψ_{in} to the imperfect optical system (cf. Fig. 1). Here we use a value $\Delta f = 10$ Å for the defocus, and employ Scherzer's condition [36] for the value of the spherical aberration, i.e., $C_s = \Delta f^2 / 1.5\lambda = 2658$ Å. Denoting complex numbers $a+ib$ by the ordered pair (a,b), this leads to $\tilde{\alpha}_{02} = \tilde{\alpha}_{20} = (-\pi\lambda\Delta f, 0) \approx (-0.7879, 0)$, and $\tilde{\alpha}_{22} = 2\tilde{\alpha}_{40} = 2\tilde{\alpha}_{04} = (\pi C_s \lambda^3, 0) \approx (0.1317, 0)$, where our calculation now leads to the values for the coefficients $\{\alpha_{mn}\}$ as

$$\alpha_{0,2} = (-0.787\,889, 0),$$

$$\alpha_{0,4} = (0.065\,865\,7, 0.310\,385),$$

$$\alpha_{0,6} = (0.081\,516\,3, -0.051\,894\,9),$$

$$\alpha_{0,8} = (-0.020\,443\,7, -0.013\,887\,3),$$

$$\alpha_{0,10} = (-0.000\,821\,096, 0.005\,369\,13),$$

$$\alpha_{0,12} = (0.001\,009\,95, -0.000\,341\,025),$$

$$\alpha_{0,14} = (-0.000\,139\,425, -0.000\,129\,127),$$

$$\alpha_{0,16} = (-7.101\,78 \times 10^{-6}, 3.036\,16 \times 10^{-5}),$$

$$\alpha_{0,18} = (4.547\,97 \times 10^{-6}, -1.419\,02 \times 10^{-6}),$$

$$\alpha_{0,20} = (-5.117\,61 \times 10^{-7}, -4.518\,82 \times 10^{-7}),$$

$$\alpha_{0,22} = (-1.719\,26 \times 10^{-8}, 9.112\,01 \times 10^{-8}),$$

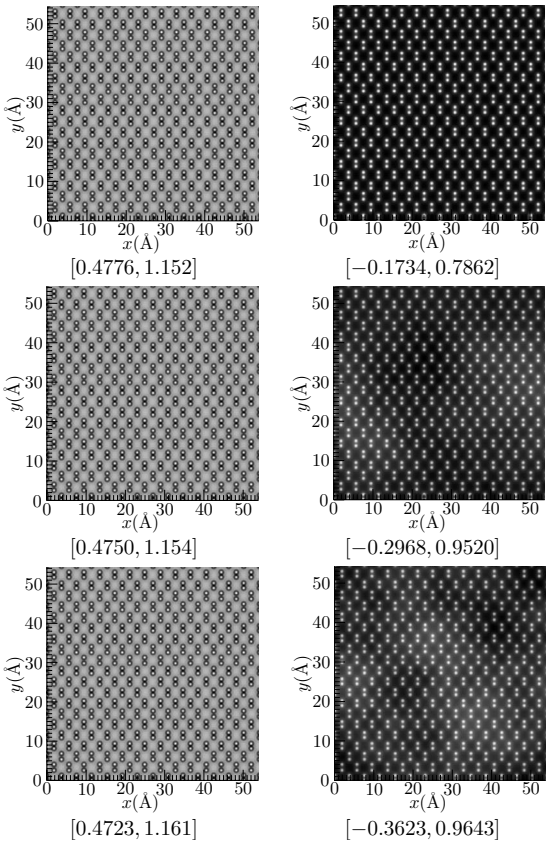

FIG. 3. Left is the output intensity for noise level 0%, 0.2%, and 0.5%, respectively, from top to bottom. The right column gives the corresponding input phase retrieved from the output intensity to its left. Below each image is the range of values corresponding to the gray level from black to white. The intensity (left) is in arbitrary units, whereas the phase (right) is in units of radians. Numbers below each image indicate the range of the gray scale value of each image. Dimensions of all images are 53.75 Å × 54.3 Å.

$$\alpha_{0,24} = (1.261\,65 \times 10^{-8}, -4.4891 \times 10^{-9}),$$

$$\alpha_{0,26} = (-1.894\,79 \times 10^{-9}, -9.388\,63 \times 10^{-10}),$$

$$\alpha_{0,28} = (1.648\,02 \times 10^{-10}, 2.253\,48 \times 10^{-10}),$$

$$\alpha_{0,30} = (-9.244\,18 \times 10^{-12}, -2.529\,66 \times 10^{-11}),$$

$$\alpha_{0,32} = (3.311\,97 \times 10^{-13}, 1.812\,06 \times 10^{-12}),$$

$$\alpha_{0,34} = (-6.921\,82 \times 10^{-15}, -8.698\,21 \times 10^{-14}),$$

$$\alpha_{0,36} = (6.429\,42 \times 10^{-17}, 2.726\,81 \times 10^{-15}),$$

$$\alpha_{0,38} = (0, -5.065\,67 \times 10^{-17}),$$

$$\alpha_{0,40} = (0, 4.234\,78 \times 10^{-19}).$$

Using the above values, we solve Eqs. (21) and (24) for the input phase using 20 iterations. Figure 3 shows the output intensity at various Poisson-noise levels (left column) and

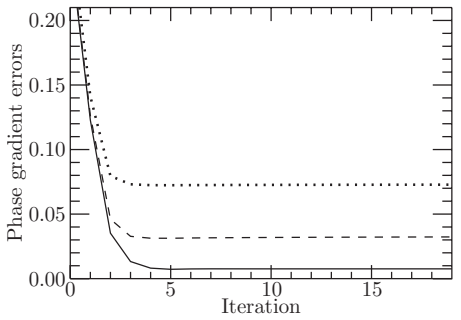

FIG. 4. Relative RMS errors in the phase gradient of the retrieved phase with various noise levels for the first 20 iterations. Solid, dashed, and dotted lines are for 0, 0.2, and 0.5 percent noise levels in the aberrated intensity, respectively.

the corresponding retrieved input phase (right column) at the end of the iterations. For our noise-free simulated data, the retrieved phase (top right of Fig. 3) is almost identical to the input phase [see Fig. 2 (right)], except for an arbitrary irrelevant global phase constant. The range of the input phases is 0.9565 radians, compared to that of the retrieved phase at 0.9596 radians. For the case of the 0.2% and 0.5% noise levels, the retrieved phase is characteristically similar to the input phase, except for the arbitrary global phase constant and some very faint low-frequency artifacts. Note that such low-frequency artifacts are a well-known characteristic of phase retrieval based on through-focal series, in the presence of noisy data [33].

The convergence and the accuracy of the retrieved phase for the various noise levels is quantitatively studied by plotting the relative root-mean-square (RMS) errors in the phase gradient

$$E = \left\{ \frac{\sum \nabla(\Phi_R - \Phi_E) \cdot \nabla(\Phi_R - \Phi_E)}{\sum \nabla\Phi_E \cdot \nabla\Phi_E} \right\}^{1/2}, \qquad (28)$$

as a function of the iteration number, where Φ_E is the exact input phase and Φ_R is the retrieved phase. The relative RMS errors are shown in Fig. 4 for various noise levels in the input and output intensities: the solid line is noise free, the dashed line corresponds to 0.2% noise, and the dotted line is for 0.5% noise. As evident from Fig. 4, the iterations converge and approach the asymptotic value in the first few iterations. For the noise-free case, the iteration starts with an error of 26%, drops down to a minimum of 0.72% at the fifth iteration, and stagnates at 0.76% as an asymptotic value. The RMS error for the case with 0.2% noise starts at 26% error, drops to a minimum of 3.1% and stagnates at 3.2%, whereas the RMS error for the case with 0.5% noise starts at 27%, drops to a minimum of 7.2%, and stagnates at 7.3% as the asymptotic value. In all numerical cases considered here, convergence is extremely rapid, being reached in five iterations or less.

IV. CONCLUSIONS AND DIRECTIONS FOR FUTURE WORK

Given input and output intensities of a complex wave field obtained from an imperfect imaging system, one faces the problem of determining the phase of the wave field. In this paper, we have successfully formulated an iteration scheme for direct solution to the phase problem given input and output intensities that are obtained using a shift-invariant coherent linear imaging system. The proposed algorithm is quite fast compared to the traditional projection method of the Gerchberg-Saxton algorithm.

The efficiency of the algorithm, which converges after a few iterations in all numerical experiments considered here, suggests that the exact analytical solution to the phase problem in optical imaging using shift-invariant linear systems may be approximated from the first iteration. The accuracy of the analytical approximation improves for smaller values of the coefficients $\{\alpha_{mn}\}$ characterizing the degree of imperfection of the system [cf. Eq. (8)]. At higher values of these coefficients, the analytical approximation is less accurate; however, it may be used as input to successive iterations or to the traditional Gerchberg-Saxton scheme for rapid convergence.

Presently our method works for the case where the intensities are obtained over the input and output planes, for a single state of the aberrated imaging system. A generalization may be possible for the case when the two intensity profiles are both taken over output planes, corresponding to output intensities obtained via two different values of the aberration coefficients, namely, two different states of the imperfect imaging system. The latter expedient may be useful if it is not possible to obtain the intensity over the input plane.

ACKNOWLEDGMENTS

We acknowledge support from the Australian Research Council (ARC), via the Discovery-Projects Program, together with useful suggestions from the anonymous referee.

[1] H. A. Hauptman, Rep. Prog. Phys. **54**, 1427 (1991).
[2] J. C. H. Spence, in *Science of Microscopy*, edited by P. W. Hawkes and J. C. H. Spence (Springer Science, LLC, 2007), Chap. 19.
[3] *Inverse Problems in Optics*, edited by H. P. Baltes (Springer-Verlag, New York, 1978).
[4] W. O. Saxton, *Computer Techniques for Image Processing in Electron Microscopy* (Academic, New York, 1978).
[5] J. R. Fienup, Appl. Opt. **21**, 2758 (1982).
[6] R. Bates, Optik (Stuttgart) **61**, 247 (1982).
[7] J. Miao, P. Charalambous, J. Kirz, and D. Sayre, Nature (London) **400**, 342 (1999).
[8] R. P. Yu, D. M. Paganin, and M. J. Morgan, Phys. Lett. A **341**, 156 (2005).
[9] R. P. Yu, D. M. Paganin, and M. J. Morgan, Phys. Rev. E **72**, 056711 (2005).

[10] R. W. Gerchberg and W. O. Saxton, Optik (Stuttgart) **35**, 237 (1972).
[11] J. R. Fienup, Opt. Lett. **3**, 27 (1978).
[12] V. Elser, J. Opt. Soc. Am. A **20**, 40 (2003).
[13] N. Nakajima, Phys. Rev. Lett. **98**, 223901 (2007).
[14] N. Nakajima, J. Opt. Soc. Am. A **25**, 742 (2008).
[15] S. G. Podorov, K. M. Pavlov, and D. M. Paganin, Opt. Express **15**, 9954 (2007).
[16] H. M. Quiney, G. J. Williams, and K. A. Nugent, Opt. Express **16**, 6896 (2008).
[17] M. Guizar-Sicairos and J. R. Fienup, Opt. Express **15**, 17592 (2007).
[18] M. Guizar-Sicairos and J. R. Fienup, Opt. Lett. **33**, 2668 (2008).
[19] A. V. Martin and L. J. Allen, Opt. Commun. **281**, 5114 (2008).
[20] I. McNulty, J. Kirz, C. Jacobsen, E. Anderson, M. Howells, and D. Kern, Science **256**, 1009 (1992).
[21] S. Eisebitt, J. Lüning, W. F. Schlotter, M. Lörgen, O. Hellwig, W. Eberhardt, and J. Stöhr, Nature (London) **432**, 885 (2004).
[22] Lorenz-M. Stadler, C. Gutt, T. Autenrieth, O. Leupold, S. Rehbein, Y. Chushkin, and G. Grubel, Phys. Rev. Lett. **100**, 245503 (2008).
[23] C. Depeursinge, in *Digital Holography and Three-Dimensional Display*, edited by T.-C. Poon (Springer, US, 2006), Chap. 4.
[24] G. Popescu, in *Methods in Cell Biology*, edited by P. J. Bhanu (Elsevier, New York, 2008), Vol. 90, Chap. 5.
[25] M. R. Teague, J. Opt. Soc. Am. **73**, 1434 (1983).
[26] D. M. Paganin and K. A. Nugent, Phys. Rev. Lett. **80**, 2586 (1998).
[27] L. J. Allen, M. P. Oxley, and D. M. Paganin, Phys. Rev. Lett. **87**, 123902 (2001).
[28] T. E. Gureyev, Y. I. Nesterets, K. M. Pavlov, and S. W. Wilkins, J. Opt. Soc. Am. A **24**, 2230 (2007).
[29] R. P. Yu and D. M. Paganin, Phys. Rev. A **78**, 055802 (2008).
[30] J. M. Cowley, *Diffraction Physics*, 3rd ed. (Elsevier Science, Amsterdam, 1995), p. 59.
[31] J. W. Goodman, *Introduction to Fourier Optics* (Roberts & Company, Greenwood Village, 2005).
[32] D. M. Paganin and T. E. Gureyev, Opt. Commun. **281**, 965 (2008).
[33] D. Paganin, A. Barty, P. J. McMahon, and K. A. Nugent, J. Microsc. **214**, 51 (2004).
[34] Y. Huang and M. A. Anastasio, J. Opt. Soc. Am. A **24**, 626 (2007).
[35] J. M. Cowley and A. F. Moodie, Acta Crystallogr. **10**, 609 (1957).
[36] E. J. Kirkland, *Advanced Computing in Electron Microscopy* (Plenum Press, New York, 1998).