Ultramicroscopy 174 (2017) 70–78

Contents lists available at ScienceDirect

Ultramicroscopy

journal homepage: www.elsevier.com/locate/ultramic

Non-iterative phase retrieval by phase modulation through a single parameter

A.P. Konijnenberg[a,*], W.M.J. Coene[a,b], H.P. Urbach[a]

[a] Optics Research Group, Delft University of Technology, Delft 2628 CH, The Netherlands
[b] ASML Netherlands B.V., De Run 6501, 5504 DR Veldhoven, The Netherlands

ABSTRACT

We report on a novel non-iterative phase retrieval method with which the complex-valued transmission function of an object can be retrieved with a non-iterative computation, with a limited number of intensity measurements. The measurements are taken in either real space or Fourier space, and for each measurement the phase in its dual space is modulated according to a single optical parameter. The requirement found for the phase modulation function is a general one, which therefore allows for plenty of customization in this method. It is shown that quantitative Zernike phase contrast imaging is one special case of this general method. With simulations we investigate the sampling requirements for a microscopy setup and for a Coherent Diffraction Imaging (CDI) setup.

1. Introduction

There are many applications where one wants to find a complex-valued function $f(x)$, but only its modulus $|f(x)|$ can be measured directly. In the context of Coherent Diffraction Imaging (CDI) this function may represent the transmission function of a sample, but there are many other applications for phase retrieval as well (e.g. quantum state tomography [1–4]). To find the function $f(x)$ itself, one must therefore find a method to retrieve the phase.

In particular, there are phase retrieval problems that involve either measurements or some kind of constraints on a Fourier transform pair, given by $f(x)$ and its transform $\tilde{f}(x') = \mathcal{F}\{f\}(x')$. An example of such a case is found in CDI. In this case we have a two-dimensional object, with a complex-valued transmission function $O(\mathbf{x})$. Here, $\mathbf{x}$ is a 2D position vector. If we illuminate the object with a plane wave we can measure the intensity of the diffraction pattern in the far field $I(\mathbf{x}') = |\tilde{O}(\mathbf{x}')|^2$, where $\tilde{O}$ denotes the Fourier transform of O, and $\mathbf{x}'$ is a 2D vector in Fourier space. Suppose, as in the original proposal by Gerchberg and Saxton [5], that we can only measure the intensity $I(\mathbf{x}')$ directly, and of the function $O(\mathbf{x})$ we only know its support (i.e. our object is an isolated object, of which we know its finite size). In other words, we have an amplitude constraint in Fourier space, and a support constraint in the object space. With projective algorithms such as the Error Reduction algorithm [5] or the Hybrid Input–Output algorithm [6], we alternatively apply the amplitude constraint and the support constraint in the two dual spaces, and that way we can try to reconstruct $O(\mathbf{x})$ and $\tilde{O}(\mathbf{x}')$. However, these algorithms are known to not always converge to the correct solution. An alternative approach is ptychography, for which algorithms have been developed such as the ptychographic iterative engine (PIE) [7]. In ptychography, an illumination function $P(\mathbf{x})$ is used to illuminate different parts of an object $O(\mathbf{x})$. That is, we shift the illumination function by some vector $\mathbf{X}_j$, and for each $\mathbf{X}_j$ we measure intensity $I_j(\mathbf{x}') = |\mathcal{F}\{O(\mathbf{x})P(\mathbf{x} - \mathbf{X}_j)\}(\mathbf{x}')|^2$. By having $P(\mathbf{x} - \mathbf{X}_j)$ overlap for different $\mathbf{X}_j$, there is redundancy in the intensity measurements $I_j(\mathbf{x}')$, which is used as an extra constraint in the reconstruction, which makes the algorithm more robust. The PIE algorithm has been extended to ePIE [8], and it has been applied to quantum tomography [4]. However, the algorithm is still a black box in the sense that there are no known guarantees for convergence to the correct solution.

The algorithms mentioned so far are all iterative methods. There are also non-iterative methods to retrieve the phase from Fourier pairs. An example of such a method is Zernike phase contrast microscopy [9]. If we have a 2D phase object $O(\mathbf{x}) = e^{i\varphi(\mathbf{x})}$, we can Fourier transform it, shift the phase of the 0th diffraction order by $\pi/2$, and apply an inverse Fourier transform. We then find that the phase information has been converted to amplitude information which can be measured directly. However, the assumption has to be made that the object is a pure phase object, and that the variation of the phase is small (i.e. the weak-phase approximation should hold). A method in which these assumptions do not have to be made is quantitative Zernike phase contrast imaging [10]. In this method, we have an arbitrary 2D complex-valued object

* Corresponding author.
E-mail address: a.p.konijnenberg@tudelft.nl (A.P. Konijnenberg).

http://dx.doi.org/10.1016/j.ultramic.2016.12.017
Received 10 October 2016; Received in revised form 23 December 2016; Accepted 24 December 2016
Available online 26 December 2016

$O(\mathbf{x})$, and we shift the phase of the 0th diffraction order of its Fourier transform $\widetilde{O}(\mathbf{x}')$ by $A_j \in [0, 2\pi)$. We then apply an inverse Fourier transform, and measure the intensity $I_j(\mathbf{x})$. By taking three different measurements for different A_j, the object $O(\mathbf{x})$ can be calculated directly. However, this approach would make it desirable that $|\widetilde{O}(\mathbf{0})|$ is sufficiently large, because otherwise the variations in $I_j(\mathbf{x})$ are too small, which makes the method very sensitive to noise.

A non-iterative phase retrieval method that in a way resembles quantitative Zernike phase contrast imaging is Fourier transform holography [11]. Whereas in the quantitative Zernike phase contrast method a perturbation (i.e. a phase-shifted pixel) is introduced inside the support of the field, in Fourier transform holography a perturbation (i.e. a point source that is coherent with the incident field) is introduced sufficiently far away from the support of the field. This way the autocorrelation of the field (which can be found by inverse Fourier transforming the intensity of the Fourier transform of the field) contains information that is proportional to the original field. The main advantage of this method is that only one intensity measurement is needed. Similar methods that use holography-related techniques with an extended reference are given in [12,13].

Another non-iterative method is the focus-variation method [14,15], for which substantial progress was made during the 1996 Brite–Euram project [16–20]. In this method, we have a 2D object $O(\mathbf{x})$, and we take intensity measurements in different defocus planes $I_j(\mathbf{x}') = |\mathcal{F}\{O(\mathbf{x})e^{iA_j|\mathbf{x}|^2}\}(\mathbf{x}')|^2$. With these intensity measurements we can directly calculate $O(\mathbf{x})$, but only in the approximation that $|\widetilde{O}(\mathbf{0})|$ is sufficiently large. If the distance between two measurement planes is sufficiently small, the Transport of Intensity Equation can be used to solve the field non-iteratively [21,22]. In this method, the difference between the intensities measured in the two planes is described with a differential equation, from which the field can be solved. A related method which uses shifting Gaussian filters is presented in [23].

A method similar to the focus-variation method is the 2D astigmatism variation method [24]. Instead of varying the defocus parameter to get different intensity measurements, two second-order astigmatism parameters are being varied. With this method, the object $O(\mathbf{x})$ can be calculated non-iteratively, and no approximation needs to be made about the magnitude of $|\widetilde{O}(\mathbf{0})|$.

An overview of various non-iterative phase retrieval methods is given in [25].

In this paper, we present another non-iterative phase retrieval method based on parameter variation. Just like in the case of focus variation and 2D astigmatism variation, we modulate the phase in one space (real space or Fourier space) according to a parameter A_j, and we measure intensities I_j in the dual space. However, as opposed to focus variation, our method does not require the approximation of $|\widetilde{O}(\mathbf{0})|$ being large, and as opposed to 2D astigmatism variation, we only need to vary one parameter. Our method gives a general form of the phase modulation function we need to apply, and we will demonstrate that in a special case this method reduces to quantitative Zernike phase contrast. Thus, in a way our general method provides a framework which connects focus variation and astigmatism variation with quantitative Zernike phase contrast, while providing an entire class of alternatives as well.

2. Method

The novel non-iterative phase retrieval method that we explain in this section can be applied in a microscopy setup (see Fig. 1a), or in a focused probe or CDI setup (see Fig. 1b). Let us for the sake of notation decide that we are treating the case for the CDI setup, but the same derivation holds for the microscopy setup if we interchange the roles of object space and Fourier space (if we assume there are no incoherent effects). It should be noted though that from a practical point of view the microscopy setup would be easier to implement than the CDI setup: in the microscopy setup one could with a Spatial Light Modulator (SLM) directly alter the phase of the field in the Fourier plane, while in the CDI setup it may not be so straightforward to shape the phase of the probe.

$O(\mathbf{x})$ can be reconstructed non-iteratively from intensity measurements as follows:

1. We have a complex-valued transmission function $O(\mathbf{x})$ of an object. We illuminate it with an illumination function $P_A(\mathbf{x}) = e^{2\pi i A f(\mathbf{x})}$.
2. For N different A, spaced by some interval Δ_A, we measure the intensity in the diffraction plane $I_A(\mathbf{x}') = |\mathcal{F}\{O \cdot P_A\}(\mathbf{x}')|^2$.
3. We reconstruct the object in $\mathbf{x} \neq \mathbf{0}$ using
$$O^*(\mathbf{0})O(\mathbf{x}) = \sum_A \mathcal{F}^{-1}\{I_A\}(\mathbf{x})H(A)e^{-2\pi i A f(\mathbf{x})}, \tag{1}$$
where $H(A)$ is a sampling function which we can choose to be e.g. Gaussian multiplied with a series of delta peaks that determine for which A we sample.
4. To reconstruct the object in $\mathbf{x} = \mathbf{0}$, we need to find $|O(\mathbf{0})|^2$. This is done by solving a quadratic equation.

In the following paragraphs we will demonstrate that this method works if $f(\mathbf{x})$ is chosen appropriately.

First, we will rewrite Eq. (1) so that it becomes more apparent why we can reconstruct $O(\mathbf{x})$ with this expression. Note that if $H(A)$ consists of multiple delta functions which indicate for which A we sample $I_A(\mathbf{x}')$, we can rewrite the right side of Eq. (1) as an integral over A
$$\sum_A \mathcal{F}^{-1}\{I_A\}(\mathbf{x})H(A)e^{-2\pi i A f(\mathbf{x})} \propto \int \mathcal{F}^{-1}\{I_A\}(\mathbf{x})H(A)e^{-2\pi i A f(\mathbf{x})}\mathrm{d}A. \tag{2}$$
We can rewrite $\mathcal{F}^{-1}\{I_A\}(\mathbf{x})$ as an autocorrelation function
$$\mathcal{F}^{-1}\{I_A\}(\mathbf{x}) = \int O(\mathbf{y})^* O(\mathbf{x}+\mathbf{y})e^{2\pi i A(f(\mathbf{x}+\mathbf{y})-f(\mathbf{y}))}\mathrm{d}\mathbf{y}. \tag{3}$$
Plugging this into Eq. (2) and defining $\widetilde{H}(A')$ as the Fourier transform of $H(A)$ we get
$$\begin{aligned}\sum_A &\mathcal{F}^{-1}\{I_A\}(\mathbf{x})H(A)e^{-2\pi i A f(\mathbf{x})}\\ &= \iint O(\mathbf{y})^* O(\mathbf{x}+\mathbf{y})e^{2\pi i A(f(\mathbf{x}+\mathbf{y})-f(\mathbf{y})-f(\mathbf{x}))}H(A)\mathrm{d}\mathbf{y}\mathrm{d}A\\ &= \int O(\mathbf{y})^* O(\mathbf{x}+\mathbf{y})\widetilde{H}(f(\mathbf{x}+\mathbf{y})-f(\mathbf{y})-f(\mathbf{x}))\mathrm{d}\mathbf{y}.\end{aligned} \tag{4}$$
Let us for now assume the ideal case that $H(A) = 1$ so that $\widetilde{H}(A') = \delta(A')$, i.e. we assume that we can sample $I_A(\mathbf{x}')$ for all A continuously. In that case Eq. (4) reduces to
$$\sum_A \mathcal{F}^{-1}\{I_A\}(\mathbf{x})H(A)e^{-2\pi i A f(\mathbf{x})} = \int O(\mathbf{y})^* O(\mathbf{x}+\mathbf{y})\delta(f(\mathbf{x}+\mathbf{y})-f(\mathbf{y}) - f(\mathbf{x}))\mathrm{d}\mathbf{y}. \tag{5}$$
Let us have a closer look at the argument of the delta function, which we define as
$$g(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}+\mathbf{y}) - f(\mathbf{y}) - f(\mathbf{x}). \tag{6}$$
Note that if $\mathbf{x} = \mathbf{0}$ or $\mathbf{y} = \mathbf{0}$, then $g(\mathbf{x}, \mathbf{y}) = 0$ (where we have assumed without loss of generality that $f(\mathbf{0}) = 0$). For now we will assume that $\mathbf{x} \neq \mathbf{0}$. Suppose that $g(\mathbf{x}, \mathbf{y}) = 0$ only if $\mathbf{y} = \mathbf{0}$. In that case Eq. (5) will reduce to
$$\sum_A \mathcal{F}^{-1}\{I_A\}(\mathbf{x})H(A)e^{-2\pi i A f(\mathbf{x})} \propto O(\mathbf{0})^* O(\mathbf{x}), \tag{7}$$
which is what we want (the expressions are in this case proportional to each other, not equal, because the determinant of the Jacobian is omitted. However, in case that we pixelate $I_A(\mathbf{x}')$ and $O(\mathbf{x})$, i.e. we discretize $\mathbf{x}$, as will always be the case in practice, this factor is irrelevant). Although the preceding derivation was not very rigorous in using delta functions, it can be made mathematically rigorous by