Optical Tomographic Image Reconstruction Based on Beam Propagation and Sparse Regularization

Ulugbek S. Kamilov, *Member, IEEE*, Ioannis N. Papadopoulos, Morteza H. Shoreh, Alexandre Goy, Cedric Vonesch, Michael Unser, *Fellow, IEEE*, and Demetri Psaltis, *Fellow, IEEE*

***Abstract*—Optical tomographic imaging requires an accurate forward model as well as regularization to mitigate missing-data artifacts and to suppress noise. Nonlinear forward models can provide more accurate interpretation of the measured data than their linear counterparts, but they generally result in computationally prohibitive reconstruction algorithms. Although sparsity-driven regularizers significantly improve the quality of reconstructed image, they further increase the computational burden of imaging. In this paper, we present a novel iterative imaging method for optical tomography that combines a nonlinear forward model based on the beam propagation method (BPM) with an edge-preserving three-dimensional (3-D) total variation (TV) regularizer. The central element of our approach is a time-reversal scheme, which allows for an efficient computation of the derivative of the transmitted wave-field with respect to the distribution of the refractive index. This time-reversal scheme together with our stochastic proximal-gradient algorithm makes it possible to optimize under a nonlinear forward model in a computationally tractable way, thus enabling a high-quality imaging of the refractive index throughout the object. We demonstrate the effectiveness of our method through several experiments on simulated and experimentally measured data.**

***Index Terms*—Optical phase tomography, total variation regularization, compressive sensing, sparse reconstruction, beam propagation method, stochastic proximal-gradient.**

I. Introduction

OPTICAL tomography is a popular and widely investigated technique for three-dimensional (3D) quantitative imaging of biological samples. In a typical setup, the sample is illuminated with a laser over multiple angles and the scattered light is holographically recorded giving access to both the amplitude and the phase of the light-field at the camera plane. The refractive index distribution of the sample is then numerically reconstructed from the scattered light-field by relying on a model describing the interaction between the sample and the light. Quantitative reconstruction of the refractive index is a central problem in biomedical imaging as it enables the visualization of the internal structure, as well as physical properties, of nearly transparent objects such as cells.

Most approaches for estimating the refractive index rely on various approximations to linearize the relationship between the refractive index and the measured light-field. For example, one approach is based on the straight-ray approximation and interprets the phase of the transmitted light-field as a line integral of the refractive index along the propagation direction. The reconstruction under straight-ray approximation can be performed efficiently by using the filtered back-projection (FBP) algorithm [1]. Another popular approach is diffraction tomography that was proposed by Wolf [2] and later refined by Devaney [3]. Diffraction tomography establishes a Fourier transform–based relationship between the measured field and the refractive index, which enables the recovery of the refractive index via a single numerical application of the inverse Fourier transform. These linear approaches are typically valid only for objects that are weakly scattering; their application on highly contrasted or large objects often results in images of poor spatial resolution.

Regularization is a standard approach for improving the resolution in optical diffraction tomography. It provides effective means for mitigating various artifacts and for suppressing noise. For example, Choi *et al.* [4] demonstrated that, under the straight-ray approximation, the missing cone artifact, which results in elongation of the reconstructed shape and underestimation of the value of the refractive index, can be significantly reduced by iteratively imposing positivity on the refractive index. The benefits of this iterative approach was further demonstrated in the weakly-scattering regime by Sung *et al.* [5]. In recent years, sparsity-promoting regularization, which is an essential component of compressive sensing theory [6], [7], has provided more dramatic improvements in the quality of tomographic imaging [8], [9]. The basic motivation is that many optical tomographic images are inherently sparse in some transform domain and can be reconstructed with high accuracy even with low amount of measured data.

In this paper, we present a novel iterative imaging method for optical tomography that combines sparsity-driven regularization with a *nonlinear* forward physical model of the propagation of the light-field. Specifically, our model is based on a

Manuscript received July 20, 2015; revised November 17, 2015; accepted January 11, 2016. Date of publication January 18, 2016; date of current version March 01, 2016. This work was supported by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC Grant 267439. This work was completed while U. S. Kamilov was with the Biomedical Imaging Group, École polytechnique fédérale de Lausanne, Lausanne, Switzerland. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Edwin A. Marengo.

U. S. Kamilov is with Mitsubishi Electric Research Laboratories, Cambridge, MA 02139 USA (e-mail: kamilov@merl.com).

I. N. Papadopoulos, M. H. Shoreh, A. Goy, and D. Psaltis are with the Optics Laboratory, École polytechnique fédérale de Lausanne, Lausanne CH-1015, Switzerland (e-mail: ioannis.papadopoulos@epfl.ch; morteza.hasanishoreh@epfl.ch; alexandre.goy@epfl.ch; demetri.psaltis@epfl.ch).

C. Vonesch and M. Unser are with the Biomedical Imaging Group, École polytechnique fédérale de Lausanne, Lausanne CH-1015, Switzerland (e-mail: cedric.vonesch@epfl.ch; michael.unser@epfl.ch).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCI.2016.2519261

2333-9403 © 2016 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

popular technique in optics called *beam propagation method* (BPM), which is extensively used for modeling diffraction and propagation effects of light-waves [10]–[14]. Accordingly, BPM provides a more accurate model than its linear counterparts, especially when scattering effects cannot be neglected. Unlike other nonlinear alternatives, such as the ones based on the coupled dipole approximation [15], [16], BPM has the advantage that it is reasonably fast to implement and that it can be efficiently optimized via a *time-reversal scheme*. This scheme allows to compute the derivative of the transmitted light field with respect to the distribution of refractive index by simple error backpropagation. This allows us to develop a fast iterative algorithm based on stochastic version of the proximal-gradient descent that uses measurements in an online fashion and thus significantly reduces the memory requirements for the reconstruction. Our results demonstrate that the proposed method can reconstruct high-quality images of the refractive index even when missing significant amounts of data.

In our companion paper, we have presented the optical and conceptual aspects of our BPM–based imaging framework [17]. Here, we complement our initial report by providing the algorithmic details of the reconstruction and by presenting additional validations on simulated as well as on experimentally measured data. Our work is also related to the recent iterative optimization method by Tian and Waller [18] that was demonstrated for imaging 3D objects using incoherent illumination and intensity detection. The primary difference is that these authors use intensity measurements directly while our method relies on digital holography [19], [20] to record the complex amplitude of the field. The other improvement is on the signal processing side with the introduction of sparse regularization in order to achieve high-quality imaging with undersampled data. An interesting future work would be to see if the method proposed in this paper works for imaging phase objects directly from their intensity measurements.

This paper is organized as follows. In Section II, we present our forward model based on BPM. In Section III, we present our algorithmic framework for the recovery of the refractive index from the measurements of the light field. Specifically, our algorithm estimates the refractive index by minimizing a cost functional, where the data-term is based on BPM and the regularizer promotes solutions with a sparse gradient. Fundamentally, the algorithm relies on the computation of the derivatives of the forward model with respect to the refractive index, which will be presented in a great detail. In Section IV, we present some experimental results illustrating the performance of our algorithm on experimental as well as simulated data.

II. Forward Model

This section presents the BPM forward model, whose complete derivation can be found in Appendix A. Although, BPM is a standard technique in optics for modeling propagation of light in inhomogeneous media [10]–[14], it is less known in the context of signal reconstruction and inverse problems. We shall denote our nonlinear forward model by $\mathbf{y} = \mathbf{S}(\mathbf{x})$, where the vector $\mathbf{y} \in \mathbb{C}^M$ contains the samples of the measured light-field, $\mathbf{x} \in \mathbb{R}^N$ is the discretized version of the refractive index, and $\mathbf{S} : \mathbb{R}^N \to \mathbb{C}^M$ is the nonlinear mapping. Note that the nonlinearity of BPM refers to the relationship between the refractive index and the measured light-field, not to the relationship between input and output light-fields, which is linear.

A. Fourier Beam-Propagation

The scalar inhomogeneous Helmholtz equation implicitly describes the relationship between the refractive index and the light field everywhere in space.

$$\left(\Delta + k^2(\boldsymbol{r})\,\mathrm{I}\right)\, u(\boldsymbol{r}) = 0, \tag{1}$$

where $\boldsymbol{r} = (x, y, z)$ denotes a spatial position, u is the total light-field at $\boldsymbol{r}$, $\Delta = (\partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2)$ is the Laplacian, I is the identity operator, and $k(\boldsymbol{r}) = \omega/c(\boldsymbol{r})$ is the wavenumber of the light field at $\boldsymbol{r}$. The spatial dependence of the wavenumber k is due to variations of the speed of light c induced by the inhomogeneous nature of the medium under consideration. Specifically, the wavenumber in (1) can be decomposed as follows

$$k(\boldsymbol{r}) = k_0 n(\boldsymbol{r}) = k_0(n_0 + \delta n(\boldsymbol{r})), \tag{2}$$

where $k_0 = \omega/c_0$ is the wavenumber in the free space, with $c_0 \approx 3 \times 10^8$ m/s being the speed of light in free space. The quantity n is the spatially varying refractive index of the sample, which we have written in terms of the refractive index of the medium n_0 and the perturbation δn due to inhomogeneities. We assume that the refractive index is real, which is an accurate approximation when imaging weakly absorbing objects such as biological cells.

BPM is a class of algorithms designed for calculating the optical field distribution in space or in time given initial conditions. By considering the complex envelope $a(\boldsymbol{r})$ of the paraxial wave $u(\boldsymbol{r}) = a(\boldsymbol{r}) \exp(\mathrm{j}k_0 n_0 z)$, one can develop BPM as an evolution equation for a in which z plays the role of evolution parameter

$$a(x, y, z + \delta z) = \mathrm{e}^{\mathrm{j}k_0(\delta n(\boldsymbol{r}))\delta z} \times \mathcal{F}^{-1}\left\{ \mathcal{F}\left\{a(\cdot, \cdot, z)\right\} \times \mathrm{e}^{-\mathrm{j}\left(\frac{\omega_x^2+\omega_y^2}{k_0 n_0 + \sqrt{k_0^2 n_0^2 - \omega_x^2 - \omega_y^2}}\right)\delta z} \right\}. \tag{3}$$

Therefore, BPM allows to obtain the wave-field in space via alternating evaluation of diffraction and refraction steps handled in the Fourier and space domains, respectively (see Appendix A for mode details).

It is important to note that BPM ignores reflections. This can be seen from the fact that if the solution exists for an arbitrary initial condition $a_0 \triangleq a(x, y, z = 0)$, then a_0 does not depend on $a(\boldsymbol{r})$.

B. Numerical Implementation

We consider a 3D volume $[-L_x/2, L_x/2] \times [-L_y/2, L_y/2] \times [0, L_z]$ that we refer to as computational

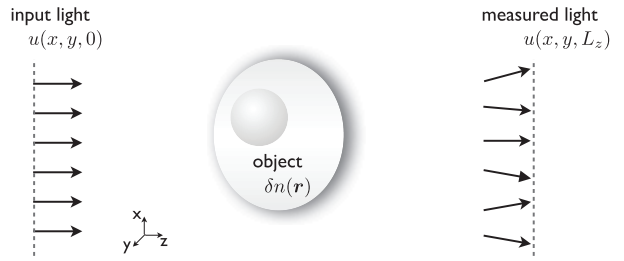

Fig. 1. Visual representation of the scattering scenario considered in this paper. A sample with a real refractive index contrast $\delta n(\boldsymbol{r})$ is illuminated with an input light $u(x, y, z = 0)$, which propagates through the sample, and results in the light $u(x, y, z = L_z)$ at the camera plane. The light at the camera plane is holographically captured and the algorithm proposed here is used for estimating the refractive index contrast $\delta n(\mathbf{r})$.

domain. The domain is sampled with steps δx, δy, and δz, which will result in N_x, N_y, and K samples, respectively. We will additionally use the following matrix-vector notations

- $\mathbf{x}$: samples of the refractive-index distribution δn in the computational domain.
- $\mathbf{y}$: samples of the complex light-field a.
- $\mathbf{S}$: nonlinear forward operator that implements BPM and maps the refractive index distribution into the complex light-field $\mathbf{y} = \mathbf{S}(\mathbf{x})$.

We use the index k to refer to the quantities described above at the kth slice along the optical axis z. For simplicity, we assume that all 2D quantities at the kth slice are stored in a vector. Then, given the initial input field $\mathbf{y}_0 = \mathbf{S}_0(\mathbf{x})$ and the refractive index distribution $\mathbf{x}$, the total field $\{\mathbf{y}_k\}_{k\in[1\ldots K]}$ can be computed recursively as follows

$$\mathbf{S}_k(\mathbf{x}) = \operatorname{diag}\left(\mathbf{p}_k(\mathbf{x}_k)\right)\mathbf{H}\mathbf{S}_{k-1}(\mathbf{x}), \tag{4}$$

where the operator $\operatorname{diag}(\mathbf{u})$ creates a square matrix with the elements of the input vector $\mathbf{u}$ the main diagonal. The matrix $\mathbf{H}$ denotes the diffraction operator; it is implemented by taking the discrete Fourier transform (DFT) of the input field, multiplying it by a frequency-domain phase mask, and taking the inverse DFT. The vector $\mathbf{p}_k(\mathbf{x}_k) = \exp(\mathrm{j}k_0\delta z\,\mathbf{x}_k)$, which depends on the kth slice of the refractive index $\mathbf{x}_k$, accounts for a phase factor corresponding to the implementation of the refraction step. Finally, the measured data $\mathbf{y}$ corresponds to the light-field at the Kth slice of the computational domain, i.e., $\mathbf{y} = \mathbf{y}_K = \mathbf{S}_K(\mathbf{x})$. Note that from (4), one can readily evaluate the computational complexity of BPM, which roughly corresponds to $2K$ evaluations of FFT or $O(N\log(N/K))$ with $N = N_xN_yK$.

Figure 2 illustrates a simulation where a plane-wave of $\lambda = 561$ nm with a Gaussian amplitude is propagated in an immersion oil ($n_0 = 1.518$ at $\lambda = 561$ nm) with an angle of $\pi/32$ with respect to the optical axis z. The computational domain of dimensions $L_x = L_y = L_z = 36.86$ μm is sampled with steps $\delta x = \delta y = \delta z = 144$ nm. In (a)–(c) we illustrate the propagation of the light-field in immersion oil, while in (d)–(f) we illustrate the propagation when a spherical bead of diameter 10 μm with refractive index $n = 1.548$ is immersed in the oil. As we can see in (f) even for a relatively weak refractive index

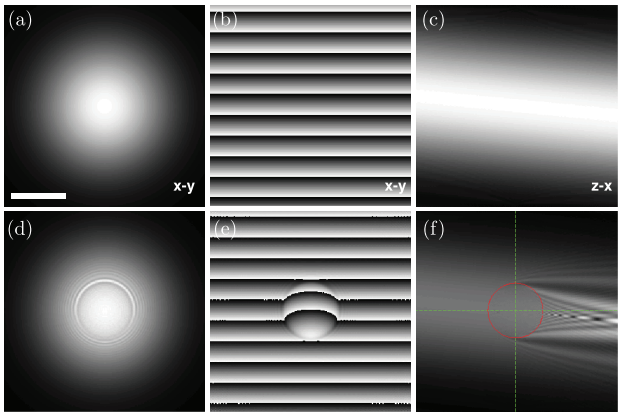

Fig. 2. Propagation of a plane-wave of $\lambda = 561$ nm in an immersion oil with $n_0 = 1.518$ simulated with BPM. (a–c) Propagation in oil. (d–f) Immersion of a 10 μm bead of $n = 1.548$. (a, d) x–y slice of the beam magnitude at $z = L_z/2$. (b, e) x–y slice of the beam phase at $z = L_z/2$. (c, f) x–z slice of the beam magnitude at $y = 0$. The circle in (f) illustrates the boundary of the bead at $y = 0$. Scale bar, 10 μm.

contrast of $\delta n = 0.03$, one can clearly observe the effects of scattering on the magnitude of the light-field.

III. Proposed Method

In practice, the input field $\mathbf{y}_0$ is known and the output field $\mathbf{y}_K$ is measured using a holographic technique that gives access to the full complex-valued light-field. Our goal is to recover $\mathbf{x}$ from a set of L views $\{\mathbf{y}_K^\ell\}_{\ell\in[1\ldots L]}$ corresponding to input fields $\{\mathbf{y}_0^\ell\}_{\ell\in[1\ldots L]}$. We shall denote with M the total number of measurements in a single view $\mathbf{y}^\ell$ and with N the total number of voxels in $\mathbf{x}$.

A. Problem Formulation

We formulate the reconstruction task as the following minimization problem

$$\widehat{\mathbf{x}} = \arg\min_{\mathbf{x}\in\mathcal{X}}\left\{\mathcal{C}(\mathbf{x})\right\} \tag{5a}$$

$$= \arg\min_{\mathbf{x}\in\mathcal{X}}\left\{\mathcal{D}(\mathbf{x}) + \tau\mathcal{R}(\mathbf{x})\right\}, \tag{5b}$$

where $\mathcal{D}$ is the data-fidelity term and $\mathcal{R}$ is the regularization term to be discussed shortly. The convex set $\mathcal{X} \subseteq \mathbb{R}^N$ is used to enforce certain physical constraints on the refractive index such as its non-negativity. The parameter $\tau > 0$ controls the amount of regularization.

The data fidelity term in (5) is given by

$$\mathcal{D}(\mathbf{x}) \triangleq \frac{1}{L}\sum_{\ell=1}^{L}\mathcal{D}_\ell(\mathbf{x}) \tag{6a}$$

$$\triangleq \frac{1}{2L}\sum_{\ell=1}^{L}\left\|\mathbf{y}_K^\ell - \mathbf{S}_K^\ell(\mathbf{x})\right\|_{\ell_2}^2, \tag{6b}$$

where L denotes the number of measured views. For a given view ℓ, the forward operator $\mathbf{S}_K^\ell$ can be computed recursively via equation (4).

Algorithm 1. Time-reversal scheme for computing $\boldsymbol{\nabla}\mathcal{D}^H$

input: input field $\mathbf{y}_0$, output field $\mathbf{y}_K$, and current estimate of the refractive-index distribution $\widehat{\mathbf{x}}$.
output: the gradient $[\boldsymbol{\nabla}\mathcal{D}(\widehat{\mathbf{x}})]^H$.
algorithm:

1) Compute the total field $\widehat{\mathbf{y}} = \mathbf{S}(\widehat{\mathbf{x}})$ using the BPM recursion (4), keeping all the intermediate light-fields $\widehat{\mathbf{y}}_k = \mathbf{S}_k(\widehat{\mathbf{x}})$ in memory.
2) Compute the residual $\mathbf{r}_K = \widehat{\mathbf{y}}_K - \mathbf{y}_K$ and set $\mathbf{s}_K = 0$.
3) Compute $\mathbf{s}_0 = \left[\frac{\partial}{\partial \mathbf{x}}\mathbf{S}_K(\widehat{\mathbf{x}})\right]^H \mathbf{r}_K$ using the following iterative procedure for $m = K, \ldots, 1$
   a) $\mathbf{s}_{m-1} = \mathbf{s}_m + \left[\frac{\partial}{\partial \mathbf{x}}\mathbf{p}_m(\widehat{\mathbf{x}}_m)\right]^H \operatorname{diag}\left(\overline{\mathbf{H}\,\widehat{\mathbf{y}}_{m-1}}\right)\mathbf{r}_m$.
   a) $\mathbf{r}_{m-1} = \mathbf{H}^H \operatorname{diag}\left(\overline{\mathbf{p}_m(\widehat{\mathbf{x}}_m)}\right)\mathbf{r}_m$.
4) Return $[\boldsymbol{\nabla}\mathcal{D}(\widehat{\mathbf{x}})]^H = \operatorname{Re}\{\mathbf{s}_0\}$.

As a regularization term in (5), we propose to use the 3D isotropic total variation (TV) [21] of the refractive index distribution

$$\mathcal{R}(\mathbf{x}) \triangleq \sum_{n=1}^{N} \|[\mathbf{D}\mathbf{x}]_n\|_{\ell_2} \tag{7}$$

$$= \sum_{n=1}^{N} \sqrt{([\mathbf{D}_x\mathbf{x}]_n)^2 + ([\mathbf{D}_y\mathbf{x}]_n)^2 + ([\mathbf{D}_z\mathbf{x}]_n)^2} \tag{8}$$

where $\mathbf{D} : \mathbb{R}^N \to \mathbb{R}^{N\times 3}$ is the discrete counterpart of the gradient operator. The matrices $\mathbf{D}_x$, $\mathbf{D}_y$, and $\mathbf{D}_z$ denote the finite difference operations along the x, y, and z directions, respectively (see Appendix B for more details). The TV prior on images has been originally introduced by Rudin *et al.* [21] as a regularization approach capable of preserving image edges, while removing noise. It is often interpreted as a sparsity-promoting ℓ_1-penalty on the magnitudes of the image gradient [22]. TV regularization has proven to be successful in a wide range of applications in the context of sparse recovery of images from incomplete or corrupted measurements [6], [23].

The minimization in (5) is a nontrivial optimization task. Keeping the regularization aside, the primary difficulty resides in the fact that the data term $\mathcal{D}$ is based on a nonlinear forward operator $\mathbf{S}$. The other challenging aspects are the massive quantity of data that need to be processed and the presence of a nonsmooth regularization term $\mathcal{R}$. We next present a novel algorithm based on iterative stochastic proximal-gradient descend that is made tractable via the time-reversal scheme that allows for an efficient computation of the gradient of $\mathcal{D}$ with respect to $\mathbf{x}$.

B. Computation of the Gradient

The crucial component of our method is recursive computation of the gradient of $\mathcal{D}$ with respect to $\mathbf{x}$, summarized in Algorithm 1, which is explained next. For notational simplicity, we consider the scenario of a single view and thus drop the indices ℓ from the subsequent derivations. The generalization of the final formula to an arbitrary number of views L is straightforward.

We start by expanding the quadratic term as

$$\mathcal{D}(\mathbf{x}) = \frac{1}{2}\|\mathbf{y}_K - \mathbf{S}_K(\mathbf{x})\|_{\ell_2}^2$$

$$= \frac{1}{2}\langle \mathbf{y}_K, \mathbf{y}_K\rangle - \operatorname{Re}\{\langle \mathbf{S}_K(\mathbf{x}), \mathbf{y}_K\rangle\} + \frac{1}{2}\langle \mathbf{S}_K(\mathbf{x}), \mathbf{S}_K(\mathbf{x})\rangle, \tag{9}$$

where $\langle \mathbf{x}, \mathbf{z}\rangle = \mathbf{z}^H\mathbf{x}$, where the Hermitian transposition ^{H} is due to the complex nature of the quantities. We adopt the convention

$$\frac{\partial}{\partial x_j}\mathbf{S}(\mathbf{x}) = \begin{bmatrix} \frac{\partial}{\partial x_j}[\mathbf{S}(\mathbf{x})]_1 \\ \vdots \\ \frac{\partial}{\partial x_j}[\mathbf{S}(\mathbf{x})]_M \end{bmatrix}. \tag{10}$$

Then, the gradient is given by

$$\boldsymbol{\nabla}\mathcal{D}(\mathbf{x}) = \left[\frac{\partial \mathcal{D}(\mathbf{x})}{\partial x_1} \cdots \frac{\partial \mathcal{D}(\mathbf{x})}{\partial x_N}\right] \tag{11}$$

$$= \operatorname{Re}\left\{(\mathbf{S}_K(\mathbf{x}) - \mathbf{y}_K)^H \left[\frac{\boldsymbol{\partial}}{\boldsymbol{\partial}\mathbf{x}}\mathbf{S}_K(\mathbf{x})\right]\right\}, \tag{12}$$

where

$$\frac{\boldsymbol{\partial}}{\boldsymbol{\partial}\mathbf{x}}\mathbf{S}_K(\mathbf{x}) = \left[\frac{\partial}{\partial x_1}[\mathbf{S}_K(\mathbf{x})] \cdots \frac{\partial}{\partial x_N}[\mathbf{S}_K(\mathbf{x})]\right]$$

$$= \begin{bmatrix} \frac{\partial}{\partial x_1}[\mathbf{S}_K(\mathbf{x})]_1 & \cdots & \frac{\partial}{\partial x_N}[\mathbf{S}_K(\mathbf{x})]_1 \\ \vdots & \vdots & \vdots \\ \frac{\partial}{\partial x_1}[\mathbf{S}_K(\mathbf{x})]_M & \cdots & \frac{\partial}{\partial x_N}[\mathbf{S}_K(\mathbf{x})]_M \end{bmatrix}.$$

In practice, we are interested in a column vector

$$[\boldsymbol{\nabla}\mathcal{D}(\mathbf{x})]^H = \operatorname{Re}\left\{\left[\frac{\boldsymbol{\partial}}{\boldsymbol{\partial}\mathbf{x}}\mathbf{S}_K(\mathbf{x})\right]^H (\mathbf{S}_K(\mathbf{x}) - \mathbf{y}_K)\right\}. \tag{13}$$

Therefore, we need to derive a tractable algorithm to compute (13). The recursive BPM formula (4) allows us to write

$$\frac{\boldsymbol{\partial}}{\boldsymbol{\partial}\mathbf{x}}\mathbf{S}_k(\mathbf{x}) = \frac{\boldsymbol{\partial}}{\boldsymbol{\partial}\mathbf{x}}\left[\operatorname{diag}(\mathbf{p}_k(\mathbf{x}_k))\,\mathbf{H}\,\mathbf{S}_{k-1}(\mathbf{x})\right]$$

$$= \operatorname{diag}(\mathbf{H}\,\mathbf{S}_{k-1}(\mathbf{x}))\left[\frac{\boldsymbol{\partial}}{\boldsymbol{\partial}\mathbf{x}}\mathbf{p}_k(\mathbf{x}_k)\right] + \operatorname{diag}(\mathbf{p}_k(\mathbf{x}_k))\,\mathbf{H}\left[\frac{\boldsymbol{\partial}}{\boldsymbol{\partial}\mathbf{x}}\mathbf{S}_{k-1}(\mathbf{x})\right].$$

Then, we have the following equality

$$\left[\frac{\boldsymbol{\partial}}{\boldsymbol{\partial}\mathbf{x}}\mathbf{S}_k(\mathbf{x})\right]^H$$

$$= \left[\frac{\boldsymbol{\partial}}{\boldsymbol{\partial}\mathbf{x}}\mathbf{S}_{k-1}(\mathbf{x})\right]^H \mathbf{H}^H \operatorname{diag}\left(\overline{\mathbf{p}_k(\mathbf{x}_k)}\right) + \left[\frac{\boldsymbol{\partial}}{\boldsymbol{\partial}\mathbf{x}}\mathbf{p}_k(\mathbf{x}_k)\right]^H \operatorname{diag}\left(\overline{\mathbf{H}\,\mathbf{S}_{k-1}(\mathbf{x})}\right), \tag{14}$$

where the vector $\overline{\mathbf{v}}$ contains complex conjugated elements of vector $\mathbf{v}$. Also, note that since the input field is known and does not depend on $\mathbf{x}$, for $k = 0$, we have

$$\left[\frac{\boldsymbol{\partial}}{\boldsymbol{\partial}\mathbf{x}}\mathbf{S}_0(\mathbf{x})\right]^H = \mathbf{0}. \tag{15}$$

Algorithm 2. Minimizes: $\mathcal{C}(\mathbf{x}) = \mathcal{D}(\mathbf{x}) + \tau\mathcal{R}(\mathbf{x})$

input: light-field data $\{\mathbf{y}^\ell\}_{\ell\in[1\dots L]}$, initial guess $\widehat{\mathbf{x}}^0$, steps $\{\gamma_t\}_{t\in\mathbb{N}}$, regularization parameter $\tau > 0$, and parameter $\tilde{L} \in [1 \dots L]$.
set: $t \leftarrow 1$, $\mathbf{s}^0 \leftarrow \widehat{\mathbf{x}}^0$, $q_0 \leftarrow 1$
repeat
Select randomly with equal probability a subset of $\tilde{L}$ views. We index them with $\{\ell_i\}_{i\in[1\dots\tilde{L}]}$
$\mathbf{z}^t \leftarrow \mathbf{s}^{t-1} - (\gamma_t/\tilde{L}) \sum_{i=1}^{\tilde{L}} \boldsymbol{\nabla}\mathcal{D}_{\ell_i}(\mathbf{s}^{t-1})$
$\widehat{\mathbf{x}}^t \leftarrow \mathrm{prox}_{\mathcal{R}}(\mathbf{z}^t, \gamma_t\tau)$
$q_t \leftarrow \frac{1}{2}\left(1 + \sqrt{1 + 4q_{t-1}^2}\right)$
$\mathbf{s}^t \leftarrow \widehat{\mathbf{x}}^t + ((q_{t-1} - 1)/q_t)(\widehat{\mathbf{x}}^t - \widehat{\mathbf{x}}^{t-1})$
$t \leftarrow t + 1$
until stopping criterion
return estimate of the refractive index $\widehat{\mathbf{x}}^t$

Based on the recursion (14) with the boundary condition (15), we obtain a practical implementation of (13), which is summarized in Algorithm 1. Conceptually, our method is similar to the *error backpropagation algorithm* extensively used in deep learning for neural networks [24]. Similarly, to backpropagation, we compute the gradient by propagating the error in a *time-reversed* fashion. Computational complexity of the time-reversal scheme is of the same order as that of BPM and essentially corresponds to a constant number of K FFTs of $N_x \times N_y$ images.

C. Iterative Reconstruction

By relying on the time-reversal scheme, we propose a novel algorithm, summarized in Algorithm 2, that reconstructs the refractive index $\mathbf{x}$ from optical tomographic measurements $\{\mathbf{y}_K^\ell\}_{\ell\in[1\dots L]}$. Conceptually, the algorithm is similar to the *fast iterative shrinkage/thresholding algorithm (FISTA)* [25], which is a popular approach for minimizing cost-functions that consist of sums between smooth and nonsmooth terms. One notable difference of Algorithm 2 with respect to FISTA, summarized in Algorithm 3 of Appendix B is that the gradient is only computed with respect to $\tilde{L} \leq L$ measurements selected with equal probability, at each iteration, from the complete set of measurements $\{\mathbf{y}_K^\ell\}_{\ell\in[1\dots L]}$. For $\tilde{L} \ll L$, this *incremental proximal-gradient* approach [26] reduces the per-iteration cost of reconstruction significantly; moreover, since gradient computation for our BPM model is highly parallelizable the number $\tilde{L}$ can be adapted to match the number of available processing units. Also, the overall convergence of Algorithm 2 can be substantially faster to that of full FISTA in Algorithm 3. To understand this consider an example where the measured views of the object are the same or very similar. Then, the partial gradient in Algorithm 2 will require $(L - \tilde{L})$ times less computation, but will still point to the right direction. A more detailed discussion on the benefits of incremental algorithms for solving very large scale optimization problems can be found in the recent work by Bertsekas [26].

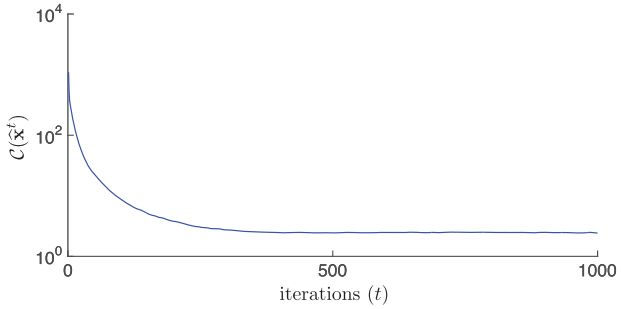

Fig. 3. Evolution of the cost $\mathcal{C}(\widehat{\mathbf{x}}^t)$ during the reconstruction over 1000 iterations for a 10 μm bead in immersion oil.

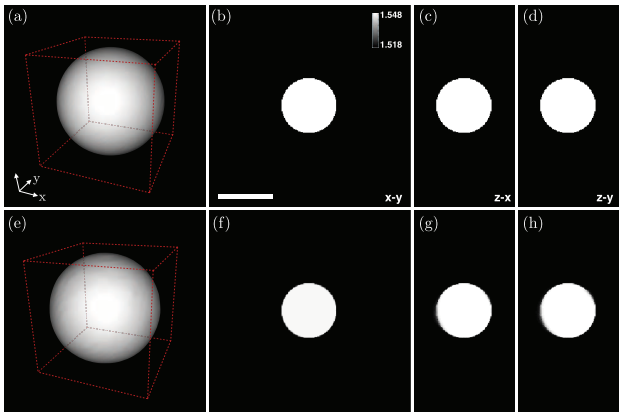

Fig. 4. Reconstruction of a 10 μm bead of refractive index 1.548 in an immersion oil with $n_0 = 1.518$ from BPM simulated measurements. (a–d) True refractive index distribution. (e–h) Reconstructed refractive index distribution: SNR = 22.74 dB. (a, e) A 3D rendered image of the bead. (b, f) x–y slice of the bead at $z = L_z/2$. (c, g) z–x slice of the bead at $y = 0$. (d, h) z–y slice of the bead at $x = 0$. Scale bar, 10 μm.

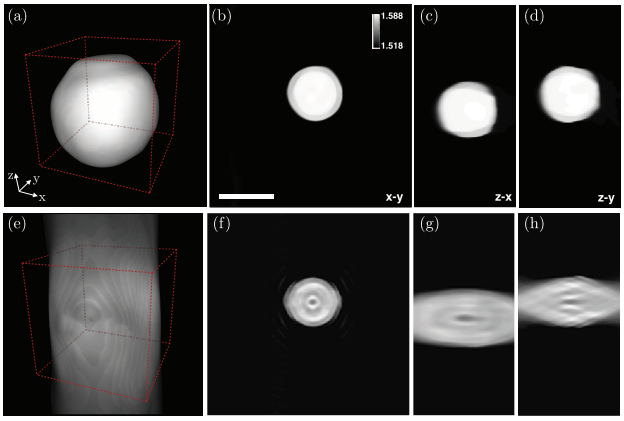

Fig. 5. Reconstruction of a 10 μm bead of refractive index 1.588 in an immersion oil with $n_0 = 1.518$ from experimentally measured data. (a–d) Reconstruction using our algorithm. (e–h) Reconstruction using the FBP algorithm. (a, e) A 3D rendered image of the bead. (b, f) x–y slice of the bead at $z = 21.17$ μm. (c, g) z–x slice of the bead at $y = -2.30$ μm. (d, h) z–y slice of the bead at $x = 0.58$ μm. Scale bar, 10 μm.

A crucial step in Algorithm 2 is the proximal operator for the regularizer $\mathcal{R}$

$$\mathrm{prox}_{\mathcal{R}}(\mathbf{z}, \tau) \triangleq \underset{\mathbf{x}\in\mathcal{X}}{\arg\min} \left\{ \frac{1}{2}\|\mathbf{x} - \mathbf{z}\|_{\ell_2}^2 + \tau\mathcal{R}(\mathbf{x}) \right\}. \quad (16)$$

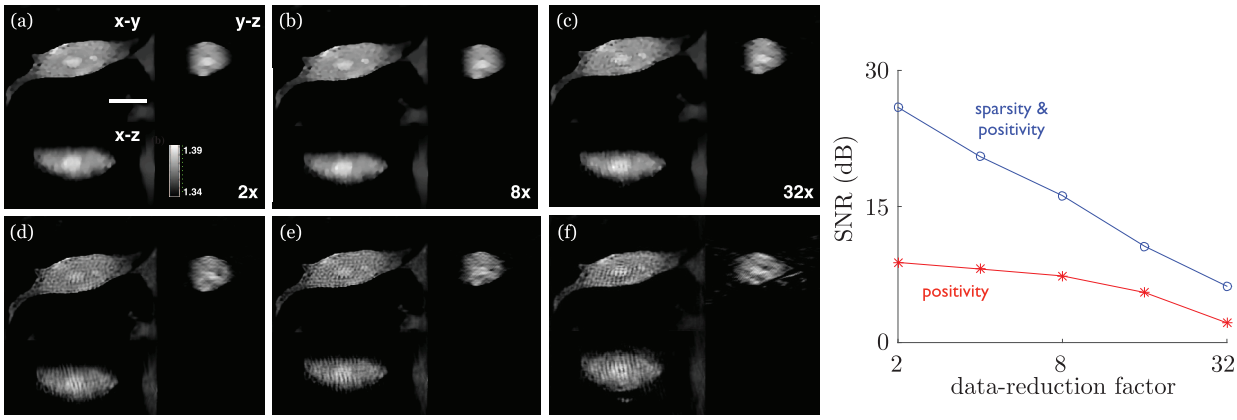

Fig. 6. Reconstruction with a proposed method of a $37 \times 37 \times 30$ μm sample containing a HeLa cell for various values of the data-reduction factor. (a–c) Reconstruction with gradient-sparsity and positivity. (d–f) Reconstruction only with positivity. (a, d) $2\times$ data reduction. (b, e) $8\times$ data reduction. (c, f) $32\times$ data reduction. Right panel shows the SNR (see text) versus the data-reduction factor for both priors. Scale bar, 10 μm.

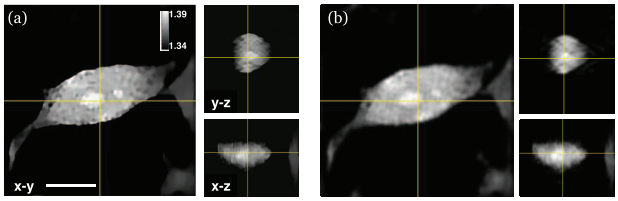

Fig. 7. Comparison of the proposed method on a HeLa cell when applying the proximal operator (a) at every iteration, (b) only once at the end for denoising purposes. The proximal operator imposes sparsity on the gradient of the image. This figure illustrates the benefits of imposing sparsity which influences the convergence to a better solution. Scale bar, 10 μm.

The proximal operator corresponds to the regularized solution of the denoising problem with the forward operator corresponding to identity. Note that although our proximal operator for 3D TV regularizer does not admit a closed form, it can be efficiently computed [25], [27]–[29]. Here, we rely on the dual minimization approach that was proposed by Beck and Teboulle [25], which we review in Appendix B and summarize in Algorithm 4.

The theoretical convergence of our algorithm is difficult to analyze due to nonlinear nature of $\mathbf{S}$. However, in practice, we found that by providing the algorithm with a warm initialization and by setting the steps of the algorithm γ_t proportional to $1/\sqrt{t}$, the algorithm achieves excellent results as reported in Section IV. The progressive reduction in γ_t is commonly done for ensuring the convergence of incremental proximal-gradient algorithms [26]. One practical approach for finding a warm initializer is to use the standard FBP algorithm that assumes a straight ray approximation. When imaging semi-transparent objects such as cells, even simpler but sufficient initialization is a constant value. Additionally, we fix the maximal number of iterations for the algorithm to $t_{\max}$ and select an additional stopping criterion based on measuring the relative change of the solution in two successive iterations as

$$\frac{\|\widehat{\mathbf{x}}^t - \widehat{\mathbf{x}}^{t-1}\|_{\ell_2}}{\|\widehat{\mathbf{x}}^{t-1}\|_{\ell_2}} \leq \delta, \tag{17}$$

where we use $\delta = 10^{-4}$ in all the experiments.

IV. Numerical Evaluation

Based on the above developments, we report the results of our iterative reconstruction algorithm in simulated and experimental configurations. The specifics of our experimental setup were discussed in the companion paper [17]. Essentially, the setup is holographic, which means that a laser source of $\lambda = 561$ nm is split into the reference and sample beams that are combined into a hologram, which is subsequently used to extract the complex light field at the measurement plane [30].

We first tested our BPM-based reconstruction algorithm on simulated data. In particular, we considered the reconstruction of a simple 10 μm bead of refractive index $n = 1.548$ immersed into oil of refractive index $n_0 = 1.518$. We simulated $L = 61$ measurements with equally spaced angles in $[-\pi/8, \pi/8]$ with BPM. The illumination beam is tilted perpendicular to the y axis, while the angle is specified with respect to the optical axis z. The dimension of computational domain is set to $L_x = L_y = 36.86$ μm and $L_z = 18.45$ μm and it is sampled with steps $\delta x = \delta y = \delta z = 144$ nm. The reconstruction is performed via the proposed approach in Algorithm 2 with $\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^N : 0 \leq \mathbf{x} \leq 0.1\}$, $\tilde{L} = 8$, and $\tau = 0.01$. In Figure 3, we illustrate the convergence of the algorithm by plotting the cost $\mathcal{C}$ for 1000 iterations. In Figure 4, we show the true and reconstructed refractive index distributions. The final signal-to-noise ratio (SNR) of the solution is 22.74 dB. The visual quality of the reconstruction is excellent; we can observe that on simulated data, the method corrects the missing cone due to limited angle of illumination and yields a sharp image along the z-axis.

We next validate the BPM forward model and our reconstruction algorithm on a similar dataset that was obtained experimentally. The sample is a 10 μm polystyrene bead of refractive index $n = 1.588$ immersed in oil with a refractive index of $n_0 = 1.518$ so that the refractive index contrast is $\delta n = 0.07$. The data was obtained by collecting $L = 61$ measurements with equally spaced angles in the range $[-32.16°, 30.80°]$. We perform reconstruction with the regularization parameter $\tau = 10$. In Figure 5(a)–(d), we show the result that was obtained by initializing our algorithm with the solution of the standard

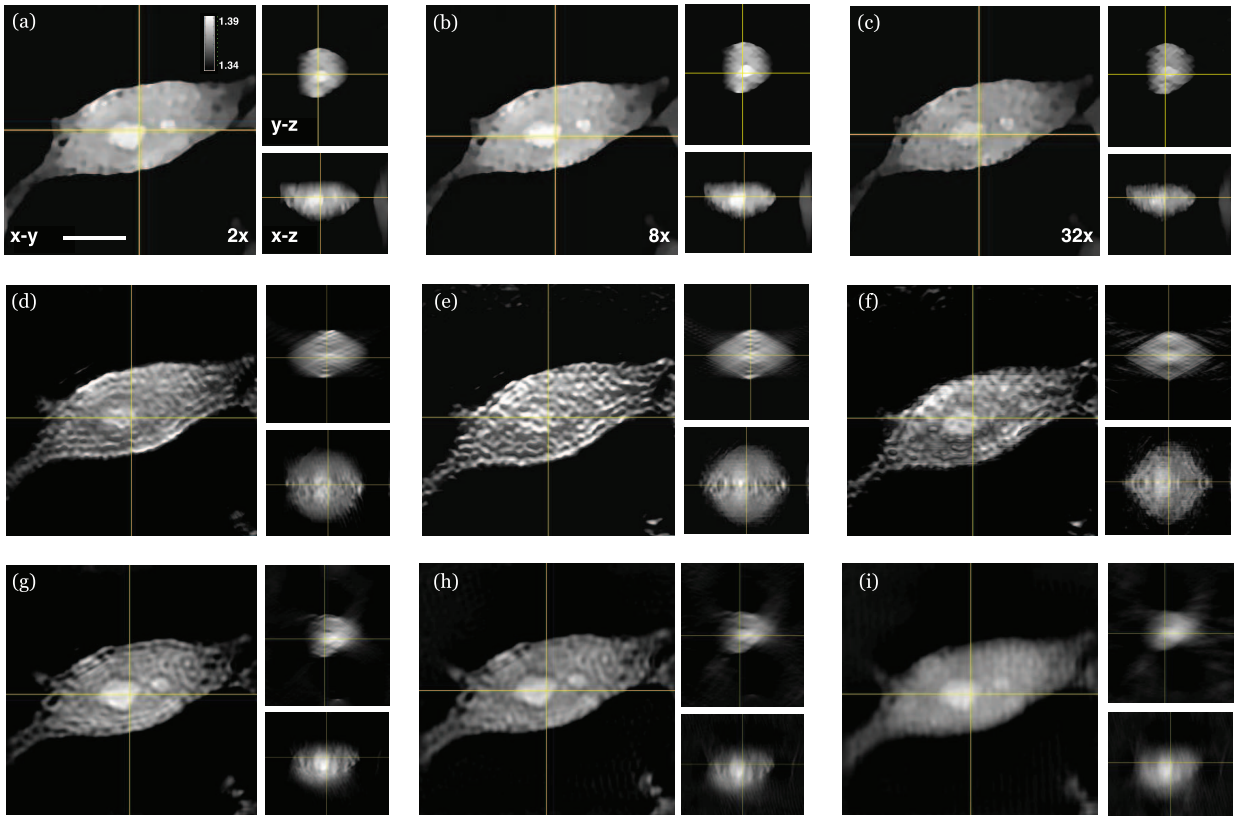

Fig. 8. Comparison of three reconstruction algorithms for various levels of data-reduction on a sample of size $37 \times 37 \times 30$ μm containing a HeLa cell. (a–c) Proposed method. (d–f) Iterative reconstruction based on a straight ray approximation [4]. (g–i) Iterative reconstruction based on diffraction tomography [31]. (a, d, g) $2\times$ data reduction. (b, e, h) $8\times$ data reduction. (c, f, i) $32\times$ data reduction. Scale bar, 10 μm.

FBP performed on the phase of the measured wave field. The FBP approach assumes a straight ray approximation and its results are illustrated in Figure 5(e)–(h). Note that such a warm initialization is useful due to the non-convex nature of our optimization problem. In the x–y slice at $z = 21.17$ μm, the bead reconstructed with our method has the diameter of approximately 10.08 μm and an average refractive index of 0.067. As we can see, one of the major benefits of using the proposed method is the correction of the missing cone that is visible in Figures 5(g) and (h).

Next, we investigated the ability of our method to reconstruct real biological samples from limited amounts of data. Specifically, we illuminated a sample containing a HeLa cell at 161 distinct angles uniformly distributed in the range $[-45°, 45°]$. The data was used for imaging a volume of size $37 \times 37 \times 30$ μm ($\delta x = \delta y = \delta z = 72$ nm). In this experiment, the data-reduction or undersampling factor refers to the ratio between the total number of holograms 161 and the actual number used for reconstruction. In particular, data-reduction factors $2, 4, 8, 16$, and 32 correspond to $81, 41, 21, 11$, and 6 holograms used for reconstruction, respectively. We illustrate the reconstruction results in Figure 6, where we compare the results of the proposed BPM–based method with and without TV regularization. We again initialize the algorithms with the volume that was obtained by running the standard FBP algorithm that assumes straight ray propagation. However, we observed that the algorithm is robust in the sense that it typically converges to the same solution independently of the initializer (also see Fig. 6 from our companion paper [17]). To quantify the quality of the reconstructed volume as a function of data-reduction factor, we also defined

$$\text{SNR (dB)} \triangleq 10 \log_{10} \left(\frac{\|\mathbf{x}_{\text{ref}}\|_{\ell_2}^2}{\|\mathbf{x}_{\text{ref}} - \widehat{\mathbf{x}}\|_{\ell_2}^2} \right),$$

where $\mathbf{x}_{\text{ref}}$ is the reconstructed volume from all the 161 possible measurements. The right panel of Figure 6 illustrates the evolution of the SNR with undersampling rate. As can be see, the sparse-regularization plays a critical role and significantly boosts the quality of the solution at all undersampling rates. Also note that the result in Figure 6(c) was obtained by using only 6 holograms of size 512×512 for reconstructing a signal of size $512 \times 512 \times 400$ voxels, which corresponds to data-to-parameter ratio of 1.5/100.

In Figure 7, we highlight the importance of sparsity-driven iterative reconstruction. Specifically, we compare our algorithm, where the TV proximal operator is applied at each iteration, against an algorithm that first reconstructs the refractive index only with positivity constraints and then applies 3D TV denoising to the final result. Although, both algorithm rely on BPM, by imposing the gradient sparsity at every iteration our algorithm converges to a visibly higher-quality solution.

In Figure 8, we compare the performance of our algorithms against two standard iterative algorithms that are commonly used in practice. The first one, whose results are shown in Figure 8 (d)–(f), is based on the algorithm that was proposed by Choi *et al.* [4]. It assumes a straight ray propagation of the light through the medium and iteratively minimizes the quadratic distance between the true and predicted phase measurements under positivity constraints. This iterative approach is an improvement over FBP and was shown to yield high quality results when imaging biological samples [4]. The second method, whose results are shown in Figure 8 (g)–(i), was proposed by Kim *et al.* [31] and is based on iterative diffraction tomography with positivity constraints. Diffraction tomography improves over the straight ray approximation by incorporating diffraction effects due to inhomogeneities in the sample into the forward model. As can be seen, our proposed method yields sharper and higher-quality images with a significant reduction in the missing cone artifacts.

V. Conclusion

We have presented a novel computational method for the estimation of the refractive index distribution of a 3D object from the measurements of the transmitted light-field. Our method relies on a nonlinear forward model, which is based on simulating the physical propagation of electromagnetic waves with BPM. We compensated the ill-posedness of the inverse problem, by imposing positivity as well as the gradient-sparsity to the solution. The method is computationally efficient due to the time-reversal scheme for computing the gradients and the fact that only a subset of gradients are evaluated at every iteration. Overall, we believe that our approach opens rich perspectives for high-resolution tomographic imaging in a range of practical setups. We have demonstrated the use of the method for experimentally reconstructing a polystyrene bead as well as a HeLa cell immersed in oil and water, respectively. Even when the number of measurements is severely restricted, the method can recover images of surprisingly high-quality.

There are several limitations that may be addressed in future work. Although, in practice, we did not encounter any convergence problems, the nonlinear nature of the forward model makes the theoretical convergence of the method difficult to analyze. Since the proposed BPM optimization scheme is similar to the error backpropagation algorithm used for training deep neural networks [32], there may be some benefit in transposing the analysis techniques that are being rapidly developed there to our framework.

In our current experimental setup the measurements are obtained by only changing the illumination angle. However, our forward model can handle arbitrary illumination patterns. This makes it much more general than its linear counterparts that are based on Radon or on diffraction tomography. Accordingly, another avenue of work would be to investigate the performance of the proposed method under different and less standard types of illumination.

Appendix A
Fourier Beam-Propagation Method

In this section, we present the full derivation that supports the use of BPM as a forward model. We start by introducing the inhomogeneous Helmholtz equation that completely characterizes the light field at all spatial positions in a time-independent form [33]. We then describe the important paraxial simplification of the Helmholtz equation, which is often used for describing the propagation of electromagnetic waves. Note that the derivations here are based on the paraxial version of BPM, which is simpler to derive, but is slightly less accurate that the nonparaxial version [34] used in (3). While, an extensive discussion on the merits and drawbacks of either version is beyond the scope of this paper, both versions are sufficiently accurate to be used in the experiments presented here.

A. Paraxial Helmholtz Equation

Our starting point is the scalar inhomogeneous Helmholtz equation

$$\left(\Delta + k^2(\boldsymbol{r})\,\mathrm{I}\right)\, u(\boldsymbol{r}) = 0,$$

where $\boldsymbol{r} = (x, y, z)$ denotes a spatial position, u is the total light-field at $\boldsymbol{r}$, $\Delta = (\partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2)$ is the Laplacian, I is the identity operator, and $k(\boldsymbol{r}) = \omega/c(\boldsymbol{r})$ is the wavenumber of the light field at $\boldsymbol{r}$. Equation (1) implicitly describes the relationship between the refractive index and the light field everywhere in space. The spatial dependence of the wavenumber k is due to variations of the speed of light c induced by the inhomogeneous nature of the medium under consideration. Specifically, the wavenumber in (1) can be decomposed as follows

$$k(\boldsymbol{r}) = k_0 n(\boldsymbol{r}) = k_0(n_0 + \delta n(\boldsymbol{r})),$$

where $k_0 = \omega/c_0$ is the wavenumber in the free space, with $c_0 \approx 3 \times 10^8$ m/s being the speed of light in free space. The quantity n is the spatially varying refractive index of the sample, which we have written in terms of the refractive index of the medium n_0 and the perturbation δn due to inhomogeneities. We next develop the paraxial Helmholtz equation for the complex envelope $a(\boldsymbol{r})$ of the paraxial wave[1]

$$u(\boldsymbol{r}) = a(\boldsymbol{r})\mathrm{e}^{\mathrm{j}k_0 n_0 z}. \tag{18}$$

One way to interpret (18) is to say that it corresponds to a plane wave propagating along z in the medium, modulated by the complex amplitude a. Now consider

$$\begin{aligned}\frac{\partial^2}{\partial z^2}u(\boldsymbol{r}) &= \frac{\partial}{\partial z}\left(\left(\frac{\partial a(\boldsymbol{r})}{\partial z}\right)\mathrm{e}^{\mathrm{j}k_0 n_0 z} + \mathrm{j}k_0 n_0 a(\boldsymbol{r})\mathrm{e}^{\mathrm{j}k_0 n_0 z}\right)\\ &= \mathrm{e}^{\mathrm{j}k_0 n_0 z}\left(\frac{\partial^2 a(\boldsymbol{r})}{\partial z^2} + 2\mathrm{j}k_0 n_0\left(\frac{\partial a(\boldsymbol{r})}{\partial z}\right) - k_0^2 n_0^2 a(\boldsymbol{r})\right).\end{aligned} \tag{19}$$

By using this expression and substituting (18) into (1), we obtain

[1]A wave is said to be paraxial if its wavefront normals are paraxial rays (i.e. when $\sin(\theta) \approx \theta$ is valid). The variation of a with position must be slow within the distance of a wavelength, so that the wave approximately maintains its underlying plane-wave nature.

$$\begin{aligned}&\left(\Delta + k^2(\boldsymbol{r})\,\mathrm{I}\right)\, u(\boldsymbol{r})\\ &= \left(\Delta_\perp + \frac{\partial^2}{\partial z^2} + 2\mathrm{j}k_0 n_0 \frac{\partial}{\partial z} - k_0^2 n_0^2 \mathrm{I} + k^2(\boldsymbol{r})\mathrm{I}\right)\\ &\qquad\qquad \times a(\boldsymbol{r})\mathrm{e}^{\mathrm{j}k_0 n_0 z}\\ &= \left(\Delta_\perp + \frac{\partial^2}{\partial z^2} + 2\mathrm{j}k_0 n_0 \frac{\partial}{\partial z} + 2k_0^2 n_0 \delta n(\boldsymbol{r})\mathrm{I} + k_0^2(\delta n(\boldsymbol{r}))^2\mathrm{I}\right)\\ &\qquad\qquad \times a(\boldsymbol{r})\mathrm{e}^{\mathrm{j}k_0 n_0 z}\\ &= 0, \end{aligned} \tag{20}$$

where $\Delta_\perp = (\partial^2/\partial x^2 + \partial^2/\partial y^2)$ is the transverse Laplacian. We now introduce two simplifications. The first is the slowly varying envelope approximation (SVEA), which is valid when $|(\partial^2/\partial z^2)a| \ll |k_0 n_0(\partial/\partial z)a|$ and which allows us to suppress the second derivative of a in z [33], [35]. In the second simplification, we ignore the term $(\delta n)^2$. We thus obtain

$$\frac{\partial}{\partial z}a(\boldsymbol{r}) = \left(\mathrm{j}\frac{1}{2k_0 n_0}\Delta_\perp + \mathrm{j}k_0\delta n(\boldsymbol{r})\mathrm{I}\right)a(\boldsymbol{r}). \tag{21}$$

Equation (21) is the slowly varying envelope approximation of the Helmholtz equation and is often referred to as the paraxial Helmholtz equation [35].

B. Fourier Beam-Propagation

BPM is a class of algorithms designed for calculating the optical field distribution in space or in time given initial conditions [14]. The paraxial Helmholtz equation (21) is an evolution equation in which the space coordinate z plays the role of evolution parameter.

We start by rewriting (21) in the operator form

$$\frac{\partial}{\partial z}a(\boldsymbol{r}) = \mathrm{D}\{a\}(\boldsymbol{r}) + \mathrm{N}\{a\}(\boldsymbol{r}), \tag{22}$$

where

$$\mathrm{D} \triangleq \mathrm{j}\frac{1}{2k_0 n_0}\Delta_\perp \quad \text{and} \quad \mathrm{N} \triangleq \mathrm{j}k_0\delta n(\boldsymbol{r})\mathrm{I}.$$

Note that the operator D is linear and translation-invariant (LTI), while the operator N corresponds to a pointwise multiplication. The solution of (22) at a sufficiently small $z = \delta z$ may be written formally as a complex exponential[2]

$$a(x, y, \delta z) = \mathrm{e}^{(\mathrm{D}+\mathrm{N})\delta z}a(x, y, 0). \tag{23}$$

The operators $\exp(\mathrm{D}z)$ and $\exp(\mathrm{N}z)$ do a priori not commute; however, Baker-Campbell-Hausdorff formula [36] can be applied to show that the error from treating them as if they do will be of order δz^2 if we are taking a small but a finite z step δz. This suggests the following approximation

$$a(x, y, z + \delta z) = \mathrm{e}^{\mathrm{N}\delta z}\mathrm{e}^{\mathrm{D}\delta z}a(x, y, z). \tag{24}$$

[2]Note that for an operator T, we define a new operator $\mathrm{e}^{\mathrm{T}z}$ in terms of series expansion $\mathrm{e}^{\mathrm{T}z} \triangleq \sum_{n=0}^{\infty}\frac{z^n}{n!}\mathrm{T}^n$. Therefore, for $a(\boldsymbol{r})$, we write $\mathrm{e}^{\mathrm{T}z}\{a\}(\boldsymbol{r}) = \sum_{n=0}^{\infty}\frac{z^n}{n!}\mathrm{T}^n\{a\}(\boldsymbol{r})$.

Now, it is possible to get explicit expressions for the diffraction $\exp(\mathrm{D}\delta z)$ and refraction $\exp(\mathrm{N}\delta z)$ operators, since they are independent. Diffraction is handled in the Fourier domain as

$$a(\omega_x, \omega_y, z + \delta z) = \mathrm{e}^{-\mathrm{j}\frac{\omega_x^2+\omega_y^2}{2k_0 n_0}\delta z}a(\omega_x, \omega_y, z), \tag{25}$$

which can also be expressed, for a fixed z, with a 2D Fourier filtering formula

$$a(x, y, z + \delta z) = \mathcal{F}^{-1}\left\{\mathcal{F}\left\{a(\cdot, \cdot, z)\right\} \times \mathrm{e}^{-\mathrm{j}\frac{\omega_x^2+\omega_y^2}{2k_0 n_0}\delta z}\right\}. \tag{26}$$

For refraction, we get

$$a(x, y, z + \delta z) = \mathrm{e}^{\mathrm{j}k_0(\delta n(\boldsymbol{r}))\delta z}a(x, y, z), \tag{27}$$

which amounts to a simple multiplication with a phase mask in the spatial domain.

A more refined version of BPM for simulating waves propagating at larger angles was derived by Feit and Flack [34]. By relying on their results, we can replace the diffraction step (26) by a more accurate alternative

$$\begin{aligned}&a(x, y, z + \delta z)\\ &= \mathcal{F}^{-1}\left\{\mathcal{F}\left\{a(\cdot, \cdot, z)\right\} \times \mathrm{e}^{-\mathrm{j}\left(\frac{\omega_x^2+\omega_y^2}{k_0 n_0 + \sqrt{k_0^2 n_0^2 - \omega_x^2 - \omega_y^2}}\right)\delta z}\right\}.\end{aligned} \tag{28}$$

Our practical implementation in Section II-B relies on this *nonparaxial* version of BPM.

APPENDIX B
TOTAL VARIATION MINIMIZATION

In this section, we discuss the concepts and algorithms behind total variation (TV) regularized image reconstruction (5). The material presented here is the review of the ideas that were originally developed by Beck and Teboulle in [25].

C. Two Variants of TV

Two common variants of TV are *anisotropic* TV regularizer

$$\mathcal{R}(\mathbf{x}) \triangleq \sum_{n=1}^{N}\|[\mathbf{D}\mathbf{x}]_n\|_{\ell_1} \tag{29a}$$

$$= \sum_{n=1}^{N}\left(|[\mathbf{D}_x\mathbf{x}]_n| + |[\mathbf{D}_y\mathbf{x}]_n| + |[\mathbf{D}_z\mathbf{x}]_n|\right) \tag{29b}$$

and *isotropic* TV regularizer

$$\mathcal{R}(\mathbf{x}) \triangleq \sum_{n=1}^{N}\|[\mathbf{D}\mathbf{x}]_n\|_{\ell_2} \tag{30a}$$

$$= \sum_{n=1}^{N}\sqrt{([\mathbf{D}_x\mathbf{x}]_n)^2 + ([\mathbf{D}_y\mathbf{x}]_n)^2 + ([\mathbf{D}_z\mathbf{x}]_n)^2} \tag{30b}$$

Algorithm 3. FISTA

input: light-field data $\{\mathbf{y}^{\ell}\}_{\ell\in[1\ldots L]}$, initial guess $\widehat{\mathbf{x}}^0$, step $\gamma > 0$, and regularization parameter $\tau > 0$.
set: $t \leftarrow 1$, $\mathbf{s}^0 \leftarrow \widehat{\mathbf{x}}^0$, $q_0 \leftarrow 1$
repeat
$\mathbf{z}^t \leftarrow \mathbf{s}^{t-1} - \gamma\boldsymbol{\nabla}\mathcal{D}(\mathbf{s}^{t-1})$
$\widehat{\mathbf{x}}^t \leftarrow \mathrm{prox}_{\mathcal{R}}(\mathbf{z}^t, \gamma\tau)$
$q_t \leftarrow \frac{1}{2}\left(1+\sqrt{1+4q_{t-1}^2}\right)$
$\mathbf{s}^t \leftarrow \widehat{\mathbf{x}}^t + ((q_{t-1}-1)/q_t)(\widehat{\mathbf{x}}^t - \widehat{\mathbf{x}}^{t-1})$
$t \leftarrow t+1$
until stopping criterion
return estimate of the refractive index $\widehat{\mathbf{x}}^t$

Here, $\mathbf{D}:\mathbb{R}^N \rightarrow \mathbb{R}^{N\times 3}$ is the discrete gradient operator, with matrices $\mathbf{D}_x$, $\mathbf{D}_y$, and $\mathbf{D}_z$ denoting the finite difference operators along x, y, and z, respectively. Assuming column-wise vectorization of a 3D matrix of size $N_y \times N_x \times N_z$, which represents the 3D image, the gradient of $\mathbf{x}$ at position $n \in [1,\ldots,N]$ is given by

$$[\mathbf{D}\mathbf{x}]_n = \begin{pmatrix}[\mathbf{D}_x\mathbf{x}]_n\\ [\mathbf{D}_y\mathbf{x}]_n\\ [\mathbf{D}_z\mathbf{x}]_n\end{pmatrix} = \begin{pmatrix}\frac{x_{n+N_y}-x_n}{\delta x}\\ \frac{x_{n+1}-x_n}{\delta y}\\ \frac{x_{n+N_xN_y}-x_n}{\delta z}\end{pmatrix}, \qquad (31)$$

with appropriate boundary conditions (periodization, Neumann boundary conditions, etc.). The constants δx, δy, and δz denote sampling intervals along x, y, and z directions, respectively. Practical implementations of TV, often assume uniform sampling by setting $\delta x = \delta y = \delta z$.

The anisotropic TV regularizer (29) can be interpreted as a sparsity-promoting ℓ_1-penalty on the image gradient, while its isotropic counterpart (30) as an ℓ_1-penalty on the *magnitudes* of the image gradient, which can also be viewed as a penalty promoting joint-sparsity of the gradient components. By promoting signals with sparse gradients, TV minimization recovers images that are piecewise-smooth, which means that they consist of smooth regions separated by sharp edges. Isotropic TV regularizer (30) is rotation invariant, which makes it preferable in the context of image reconstruction.

One must note that similar to other regularization schemes, there is, unfortunately, no theoretically optimal way of setting τ; its optimal value might depend on a number of parameters including the sample, forward model, and noise. Generally, higher levels of τ imply stronger regularization during the reconstruction and the optimal value of τ, in our experiments, was in the range $[10^{-2}, 10^{1}]$ for the configurations considered.

B. Minimization of TV

Fast iterative shrinkage/thresholding algorithm (FISTA), summarized in Algorithm 3, is one of the most popular approaches for solving (5). FISTA relies on the efficient evaluation of the gradient $\boldsymbol{\nabla}\mathcal{D}$ and of the proximal operator (16). Time-reversal scheme, in Algorithm 1, makes application of FISTA straightforward for solving (5) with regularizers that admit closed form poximal operators such as ℓ_1-penalty.

Algorithm 4. FGP for evaluating $\mathbf{x} = \mathrm{prox}_{\mathcal{R}}(\mathbf{z},\tau)$

input: $\mathbf{z} \in \mathbb{R}^N$, $\tau > 0$.
set: $t \leftarrow 1$, $\mathbf{d}^0 \leftarrow \mathbf{g}^0$, $q_0 \leftarrow 1$, $\gamma \leftarrow 1/(12\tau)$
repeat
$\mathbf{g}^t \leftarrow \mathrm{proj}_{\mathcal{G}}\left(\mathbf{d}^{t-1} + \gamma\mathbf{D}\left(\mathrm{proj}_{\mathcal{X}}\left(\mathbf{z} - \tau\mathbf{D}^T\mathbf{d}^{t-1}\right)\right)\right)$
$\mathbf{x}^t = \mathrm{proj}_{\mathcal{X}}\left(\mathbf{z} - \tau\mathbf{D}^T\mathbf{g}^t\right)$
$q_t \leftarrow \frac{1}{2}\left(1+\sqrt{1+4q_{t-1}^2}\right)$
$\mathbf{d}^t \leftarrow \mathbf{g}^t + ((q_{t-1}-1)/q_t)(\mathbf{g}^t - \mathbf{g}^{t-1})$
$t \leftarrow t+1$
until stopping criterion
return $\mathbf{x}^t$

However, some regularizers including TV do not have closed form proximals and require an additional iterative algorithm for solving (16).

In our implementation, we solve (16) with the dual approach that was proposed by Beck and Teboulle in [25]. The approach, summarized in Algorithm 4, is based on iterative solving the dual optimization problem

$$\hat{\mathbf{g}} = \underset{\mathbf{g}\in\mathcal{G}}{\arg\min}\left\{\mathcal{Q}(\mathbf{g})\right\}, \qquad (32)$$

where

$$\mathcal{Q}(\mathbf{g}) \triangleq -\frac{1}{2}\|\mathbf{z} - \tau\mathbf{D}^T\mathbf{g} - \mathrm{proj}_{\mathcal{X}}(\mathbf{z} - \tau\mathbf{D}^T\mathbf{g})\|_{\ell_2}^2 \qquad (33)$$

$$+\frac{1}{2}\|\mathbf{z} - \tau\mathbf{D}^T\mathbf{g}\|_{\ell_2}^2. \qquad (34)$$

Given the dual iterate $\mathbf{g}^t$, the corresponding primal iterate can be computed as

$$\mathbf{x}^t = \mathrm{proj}_{\mathcal{X}}(\mathbf{z} - \tau\mathbf{D}^T\mathbf{g}^t). \qquad (35)$$

The operator $\mathrm{proj}_{\mathcal{X}}$ represents an orthogonal projection onto the convex set $\mathcal{X}$. For example, a projection onto N-dimensional cube

$$\mathcal{X} \triangleq \left\{\mathbf{x}\in\mathbb{R}^N \,:\, a \le x_n \le b,\, \forall n \in [1,\ldots,N]\right\}, \qquad (36)$$

with bounds a, $b > 0$, is given by

$$[\mathrm{proj}_{\mathcal{X}}(\mathbf{x})]_n = \begin{cases} a & \text{if } x_n < a\\ x_n & \text{if } a \le x_n \le b\\ b & \text{if } x_n > b,\end{cases} \qquad (37)$$

for all $n \in [1,\ldots,N]$.

The set $\mathcal{G} \subseteq \mathbb{R}^{N\times 3}$ in (32) depends on the variant of TV used for regularization. For anisotropic TV (29), the set corresponds to

$$\mathcal{G} \triangleq \{\mathbf{g}\in\mathbb{R}^{N\times 3} : \|[\mathbf{g}]_n\|_{\ell_\infty} \le 1, \forall n \in [1,\ldots,N]\} \qquad (38)$$

with the corresponding projection

$$[\mathrm{proj}_{\mathcal{G}}(\mathbf{g})]_n = \begin{pmatrix}\frac{[\mathbf{g}_x]_n}{\max(1,|[\mathbf{g}_x]_n|)}\\ \frac{[\mathbf{g}_y]_n}{\max(1,|[\mathbf{g}_y]_n|)}\\ \frac{[\mathbf{g}_z]_n}{\max(1,|[\mathbf{g}_z]_n|)}\end{pmatrix}, \qquad (39)$$

for all $n \in [1, \ldots, N]$. Similarly, for isotropic TV (30), the set corresponds to

$$\mathcal{G} \triangleq \{\mathbf{g} \in \mathbb{R}^{N \times 3} : \|[\mathbf{g}]_n\|_{\ell_2} \leq 1, \forall n \in [1, \ldots, N]\} \quad (40)$$

with the corresponding projection

$$[\mathrm{proj}_{\mathcal{G}}(\mathbf{g})]_n = \frac{[\mathbf{g}]_n}{\max\left(1, \|[\mathbf{g}]_n\|_{\ell_2}\right)}, \quad (41)$$

for all $n \in [1, \ldots, N]$.

While the theoretical convergence of FISTA requires the full convergence of inner Algorithm 4, in practice, it is sufficient to run about 5-10 iterations with an initializer that corresponds to the dual variable from the previous outer iteration. In our implementation, we thus fix the maximal number of inner iterations to $t_{\text{in}} = 10$ and enforce an additional stopping criterion based on measuring the relative change of the solution in two successive iterations as $\|\mathbf{g}^t - \mathbf{g}^{t-1}\|_{\ell_2} / \|\mathbf{g}^{t-1}\|_{\ell_2} \leq \delta_{\text{in}}$, where $\delta_{\text{in}} = 10^{-4}$ in all the experiments here.

REFERENCES

[1] A. C. Kak and M. Slaney, *Principles of Computerized Tomography Imaging*. Piscataway, NJ, USA: IEEE, 1988.

[2] E. Wolf, "Three-dimensional structure determination of semi-transparent objects from holographic data," *Opt. Commun.*, vol. 1, no. 4, pp. 153–156, Sep./Oct. 1969.

[3] A. J. Devaney, "Inverse-scattering theory within the Rytov approximation," *Opt. Lett.*, vol. 6, no. 8, pp. 374–376, Aug. 1981.

[4] W. Choi *et al.*, "Tomographic phase microscopy," *Nat. Methods*, vol. 4, no. 9, pp. 717–719, Sep. 2007.

[5] Y. Sung, W. Choi, C. Fang-Yen, K. Badizadegan, R. R. Dasari, and M. S. Feld, "Optical diffraction tomography for high resolution live cell imaging," *Opt. Express*, vol. 17, no. 1, pp. 266–277, Dec. 2009.

[6] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[7] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[8] M. M. Bronstein, A. M. Bronstein, M. Zibulevsky, and H. Azhari, "Reconstruction in diffraction ultrasound tomography using nonuniform FFT," *IEEE Trans. Med. Imag.*, vol. 21, no. 11, pp. 1395–1401, Nov. 2002.

[9] Y. Sung and R. R. Dasari, "Deterministic regularization of three-dimensional optical diffraction tomography," *J. Opt. Soc. Amer. A*, vol. 28, no. 8, pp. 1554–1561, Aug. 2011.

[10] H. J. Breaux, "An analysis of mathematical transformations and a comparison of numerical techniques for computation of high-energy CW laser propagation in an inhomogeneous medium," Ballistic Research Laboratories Aberdeen Proving Ground, Aberdeen, MD, USA, Tech. Rep. BRL-1723, Jun. 1974.

[11] J. Wallace and J. Q. Lilly, "Thermal blooming of repetitively pulsed laser beams," *J. Opt. Soc. Amer.*, vol. 64, no. 12, pp. 1651–1655, 1974.

[12] J. A. Fleck, J. R. Morris, and M. D. Feit, "Time-dependent propagation of high energy laser beams through the atmosphere," *Appl. Phys.*, vol. 10, no. 2, pp. 129–160, Jun. 1976.

[13] W. G. Tam, "Split-step Fourier-transform analysis for laser-pulse propagation in particulate media," *J. Opt. Soc. Amer.*, vol. 72, no. 10, pp. 1311–1316, Oct. 1982.

[14] A. Goy, "Imaging and microscopy in linear and nonlinear media using digital holography," Ph.D. dissertation, School Eng., École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, Jan. 2013.

[15] G. Maire et al., "Experimental demonstration of quantitative imaging beyond Abbe's limit with optical diffraction tomography," *Phys. Rev. Lett.*, vol. 102, p. 213905, May 2009.

[16] O. Haeberlé, K. Belkebir, H. Giovaninni, and A. Sentenac, "Tomographic diffractive microscopy: Basic, techniques, and perspectives," *J. Mod. Opt.*, vol. 57, no. 9, pp. 686–699, May 2010.

[17] U. S. Kamilov *et al.*, "Learning approach to optical tomography," *Optica*, vol. 2, no. 6, pp. 517–522, Jun. 2015.

[18] L. Tian and L. Waller, "3D intensity and phase imaging from light field measurements in an LED array microscope," *Optica*, vol. 2, pp. 104–111, 2015.

[19] I. Yamaguchi and T. Zhang, "Phase-shifting digital horography," *Opt. Lett.*, vol. 22, pp. 1268–1270, 1997.

[20] U. Schnars and W. Jueptner, *Digital Horography*. New York, NY, USA: Springer, 2005.

[21] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, nos. 1–4, pp. 259–268, Nov. 1992.

[22] M. Unser and P. Tafti, *An Introduction to Sparse Stochastic Processes*. Cambridge, U.K.: Cambridge Univ. Press, 2014.

[23] E. Bostan, U. S. Kamilov, M. Nilchian, and M. Unser, "Sparse stochastic processes and discretization of linear inverse problems," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2699–2710, Jul. 2013.

[24] C. M. Bishop, *Neural Networks for Pattern Recognition*. London, U.K.: Oxford Univ. Press, 1995.

[25] A. Beck and M. Teboulle, "Fast gradient-based algorithm for constrained total variation image denoising and deblurring problems," *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2419–2434, Nov. 2009.

[26] D. P. Bertsekas, "Incremental proximal methods for large scale convex optimization," *Math. Program. Ser. B*, vol. 129, pp. 163–195, 2011.

[27] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2345–2356, Sep. 2010.

[28] U. S. Kamilov, E. Bostan, and M. Unser, "Variational justification of cycle spinning for wavelet-based solutions of inverse problems," *IEEE Signal Process. Lett.*, vol. 21, no. 11, pp. 1326–1330, Nov. 2014.

[29] Z. Qin, D. Goldfarb, and S. Ma, "An alternating direction method for total variation denoising," *Optim. Method Softw.*, vol. 30, no. 3, pp. 594–615, 2015.

[30] A. Bourquard, N. Pavillon, E. Bostan, C. Depeursinge, and M. Unser, "A practical inverse-problem approach to digital holographic reconstruction," *Opt. Express*, vol. 21, no. 3, pp. 3417–3433, Feb. 2013.

[31] T. Kim *et al.*, "White-light diffraction tomography of unlabelled live cells," *Nat. Photonics*, vol. 8, pp. 256–263, Mar. 2014.

[32] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

[33] J. W. Goodman, *Introduction to Fourier Optics*, 2nd ed. New York, NY, USA: McGraw-Hill, 1996.

[34] M. D. Feit and J. A. Fleck, "Bean nonparaxiality, filament formaform, and beam breakup in the self-focusing of optical beams," *J. Opt. Soc. Amer. B*, vol. 5, no. 3, pp. 633–640, Mar. 1988.

[35] B. Saleh and M. Teich, *Fundamentals of Photonics*, 2nd ed. Hoboken, NJ, USA: Wiley, 2007.

[36] R. Gilmore, "Baker-Campbell-Hausdorff formulas," *J. Math. Phys.*, vol. 15, pp. 2090–2092, 1974.

Ulugbek S. Kamilov (S'11–M'15) received the Ph.D. degree from the Department of Electrical Engineering, École polytechnique fédérale de Lausanne (EPFL), Lausanne, Switzerland, in 2015. Since 2015, he has been a Research Staff Member with Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA, USA. His research develops computational techniques for solving inverse problems for applications in biomedical or industrial imaging. Prior to joining MERL, he was an Exchange Student with Carnegie Mellon University, Pittsburgh, PA, USA, in 2007, a Visiting Student at Massachusetts Institute of Technology, Cambridge, MA, USA, in 2010, and a Visiting Student Researcher at Stanford University, Stanford, CA, USA, in 2013. His research interests include image processing, statistical estimation, signal representation, and machine learning.

Ioannis N. Papadopoulos received the Diploma degree in electrical and computer engineering from the Aristotle University of Thessaloniki, Thessaloniki, Greece, and the Ph.D. degree from the Laboratory of Optics, École polytechnique fédérale de Lausanne, Lausanne, Switzerland, in 2009 and 2014, respectively. He is currently working as a Research Fellow with the Bioimaging and Neurophotonics Laboratory, Neurocure Cluster of Excellence, Charite University Hospital, Berlin, Germany. His research interests include biomedical imaging, imaging through scattering media, multimode fiber endoscopy, and optical tomography. He was the recipient of the Early Post-Doc Mobility Fellowship of the Swiss National Science Foundation.

Morteza H. Shoreh was born in Tehran, Iran, on January 9, 1988. He received the B.Sc. degree in electrical engineering from the University of Tehran (UT), Tehran, Iran, and the M.Sc. degree in communication system from the Department of Electrical Engineering, Sharif University of Technology (SUT), Tehran, Iran, in 2010 and 2012, respectively. He is currently pursuing the Ph.D. degree at the Optics Laboratory (LO), École polytechnique fédérale de Lausanne (EPFL), Lausanne, Switzerland. He was with Optical Networks Research Laboratory (ONRL), from 2011 to 2015. Since 2015, he has been with Optics Laboratory (LO), École polytechnique fédérale de Lausanne (EPFL). His research interests include nonlinear optics, super-resolution, machine learning algorithms, biological imaging, and bio-photonics.

Alexandre Goy received the Master's degree in micro-engineering from École polytechnique fédérale de Lausanne (EPFL), Lausanne, Switzerland, the Master's degree in geology from Lausanne University, Lausanne, Switzerland, the Ph.D. degree in nonlinear optics from EPFL, in 2003, 2008, and 2013, respectively. He also worked on confocal microscopy and digital holography for three-dimensional cell imaging. He went for two-year postdoc to Princeton University, New Jersey, where he worked on photoacoustic imaging, nonlinear digital imaging, and ultrasound probing for batteries. He is currently working in the laboratory of Prof. Demetri Psaltis at EPFL, where he works on optical inverse scattering and microscopy.

Cédric Vonesch was born on April 14, 1981, in Obernai, France. He received the M.S. degree from École polytechnique fédérale de Lausanne (EPFL), Lausanne, Switzerland, and the Ph.D. degree from EPFL's Biomedical Imaging Group, in 2004 and 2009, respectively. He then joined the Program in Applied and Computational Mathematics at Princeton University for a postdoctoral position. From 2012 to 2015, he was again with EPFL's Biomedical Imaging Group. His research interests include inverse problems in the imaging sciences, multiresolution (wavelet-based) methods, and applications to fluorescence and electron microscopy, as well as seismic imaging.

Michael Unser (M'89–SM'94–F'99) is a Professor and the Director of EPFL's Biomedical Imaging Group, Lausanne, Switzerland. From 1985 to 1997, he was with the Biomedical Engineering and Instrumentation Program, National Institutes of Health, Bethesda, MD, USA, conducting research on bioimaging. His research interests include biomedical image processing. He is internationally recognized for his research contributions to sampling theory, wavelets, the use of splines for image processing, stochastic processes, and computational bioimaging. He has authored over 250 journal papers on those topics. He has authored (with P. Tafti) the book *An Introduction to Sparse Stochastic Processes* (Cambridge University Press, 2014). He has held the position of Associate Editor-in-Chief (2003–2005) for the IEEE TRANSACTIONS ON MEDICAL IMAGING. He is currently member of the Editorial Boards of *SIAM Journal of Imaging Sciences*, the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, and *Foundations and Trends in Signal Processing*. He is the founding Chair of the technical committee on Bio Imaging and Signal Processing (BISP) of the IEEE Signal Processing Society. He is an EURASIP Fellow (2009), and a Member of the Swiss Academy of Engineering Sciences. He was the recipient of several international prizes including three IEEE-SPS best paper awards and two technical achievement awards from the IEEE (2008 SPS and EMBS 2010).

Demetri Psaltis (F'05) received the degree from Carnegie-Mellon University, Pittsburgh, PA, USA. He is a Professor of Optics, the Director of the Optics Laboratory, and the Dean of the Engineering Schools, École polytechnique fédérale de Lausanne (EPFL), Lausanne, Switzerland. In 1980, he joined the Faculty at the California Institute of Technology, Pasadena, CA, USA. He moved to EPFL in 2006. His research interests include imaging, holography, biophotonics, nonlinear optics, and optofluidics. He has authored or coauthored over 400 publications in these areas. He is a Fellow of the Optical Society of America, the European Optical Society, and the Society for Photo-optical Systems Engineering (SPIE). He was the recipient of the International Commission of Optics Prize, the Humboldt Award, the Leith Medal, and the Gabor Prize.