

# Phase retrieval in $4f$ optical system: background compensation and sparse regularization of object with binary amplitude

Artem Migukin,<sup>1,\*</sup> Mostafa Agour,<sup>2,3</sup> and Vladimir Katkovnik<sup>1</sup>

<sup>1</sup>Department of Signal Processing, Tampere University of Technology, P.O. Box 527, Tampere FI-33101, Finland

<sup>2</sup>Bremer Institut für angewandte Strahltechnik, Klagenfurter Str. 2, Bremen D-28359, Germany

<sup>3</sup>Physics Department, Faculty of Science, Aswan University, Aswan 81528, Egypt

\*Corresponding author: artem.migukin@tut.fi

Received 23 July 2012; revised 18 October 2012; accepted 18 October 2012;  
posted 19 October 2012 (Doc. ID 173116); published 28 November 2012

Generally, wave field reconstructions obtained by phase-retrieval algorithms are noisy, blurred, and corrupted by various artifacts such as irregular waves, spots, etc. These distortions, arising due to many factors, such as nonidealities of the optical system (misalignment, focusing errors), dust on optical elements, reflections, and vibration, are hard to localize and specify. It is assumed that there is a cumulative disturbance called “background,” which describes mentioned distortions in the coherent imaging system manifested at the sensor plane. Here we propose a novel iterative phase-retrieval algorithm compensating for these distortions in the optical system. An estimate of this background is obtained via special calibration experiments, and then it is used for the object reconstruction. The algorithm is based on the maximum likelihood approach targeting on the optimal object reconstruction from noisy data and imaging enhancement using *a priori* information on the object amplitude. In this work we demonstrate the compensation of the distortions of the optical trace for a complex-valued object with a binary amplitude. The developed algorithm results in state-of-the-art filtering, and sharp reconstruction imaging of the object amplitude can be achieved. © 2012 Optical Society of America

*OCIS codes:* 030.4280, 050.1960, 070.2025, 100.3010, 100.3190, 100.5070.

## 1. Introduction

The conventional sensors detect only the intensity of the light. The phase of the light wave field contains, however, important information on the object, which is useful in, e.g., microscopy, astronomy, and material analysis. Since the phase cannot be recorded directly and it is systematically lost in the physical measurements, computational phase-recovering techniques are required for imaging and data processing. Phase recovering and, in general, the reconstruction of the object amplitude and phase is referred to as the phase-retrieval problem.

Perhaps the first iterative method for phase retrieval from intensity measurements was the well-known Gerchberg–Saxton algorithm [1], initially employed for a single observation plane, and its variation devised by Misell [2] for two defocusing images at different measurement planes. The idea consisting in the iterative replacement of the estimated magnitude by measured and prior information was further developed for various applications by many authors (e.g., [3–6]). Similar methods are proposed for Fresnel instead of Fourier transforms as the transfer functions of the wave field propagation [7–9]. Various phase-retrieval algorithms based on these landmark works are systematized by Fienup [10], who introduced classical types of iterative phase-retrieval algorithms. Multiple measurements gain an observation

redundancy that can be exploited in order to improve the quality of the complex-valued object reconstruction [11–13].

The above-mentioned imaging techniques are mainly based on an *ideal* wave field propagation modeling derived from the scalar diffraction theory [14]. In practice, wave fields in real coherent imaging systems and their observations are quite different from those predicted by theory; hence wave field reconstructions obtained by simulations (i.e., theoretical results) and using real experimental data can dramatically vary. The reconstructions obtained from the real data differ from simulated ones by multiple well-seen artifacts in the form of irregular waves, spots, random noise, etc. These systematic distortions appear due to many factors such as nonidealities of the optical system (misalignment, focusing errors, aberrations), dust on optical elements, reflections, vibration, etc.

In this paper, we consider a  $4f$  optical system with a spatial light modulator (SLM) located across the Fourier domain of the first lens. This system, imitating the lensless optical setup for the free-space diffraction propagation [15], is used for capturing multiple intensity observations at the sensor plane for the phase reconstruction (see Fig. 1). The reconstruction from these data varies from the theoretical prediction due to the different propagation operator of the used optical mask realized by a phase-modulating SLM. Moreover, the used SLM is one of the strongest sources of disturbances of the optical system due to its location at the Fourier plane.

In general, there is a diversity of numerical approaches, which are used for calibration [16], filtering parasitic reflections [17], compensating for curvature introduced by microscope objective [18,19], aberrations [20], or astigmatism [21]. In this work we assume that there is a *generalized function* at the object plane that describes various distortions in the coherent imaging system, manifested at the sensor plane, and errors of the numerical model. Namely, we do not try to identify particular sources

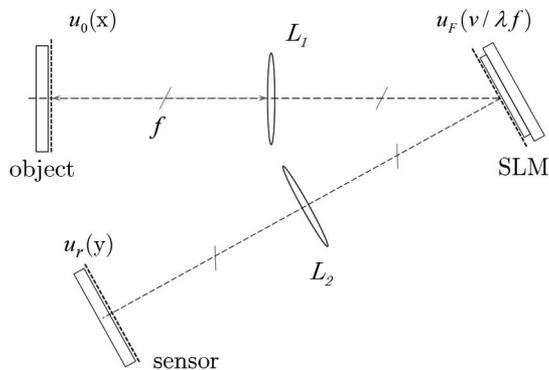


Fig. 1. Experimental setup of the  $4f$  optical system used for recording measurement data [15]. Lenses  $L_1$  and  $L_2$  in the  $4f$  configuration provide an accurate mapping of the object wave field to the parallel observation (sensor) plane. An optical mask with the complex-valued transmittance  $M_r$  located at the Fourier plane (a phase-modulating SLM) enables linear filter operations.

of the disturbances but estimate and compensate for their accumulated effects by recalculating them to the entrance pupil of the used  $4f$  configuration. In the following, the cumulative distortions are referred to as “background” distortions. Thus, we are developing an iterative phase-retrieval algorithm, where, first, we estimate this background disturbance using special calibration experiments and then use it to reconstruct the object amplitude and phase. In this work we apply the variational constrained maximum likelihood formulation with parallel processing of multiple intensity observations proposed in our previous works [22–24].

In this work we demonstrate the performance and imaging enhancement of the proposed phase-retrieval algorithm with the background compensation for a complex-valued object with a binary amplitude. We incorporate prior information on the structure of the true object wave field, and reconstruct both unknown lower and upper levels of the binary amplitude and find indices of these levels.

Let  $u_0(x)$ ,  $x \in R^2$  be a true complex-valued object at the entrance pupil of the optical system. Modeling the nonideality of the optical system, we introduce a “disturbed” object  $\tilde{u}_0(x)$  as a product of a typically unknown background (cumulative complex-valued disturbance)  $u_B(x)$  by the true object  $u_0(x)$  as

$$\tilde{u}_0(x) = u_0(x) \cdot u_B(x), \quad (1)$$

where the diacritic  $\sim$  emphasizes the difference of the disturbed  $\tilde{u}_0$  from the true/ideal one  $u_0$ . The standard phase-retrieval techniques are able to give the reconstruction of  $\tilde{u}_0(x)$  only and they are not able to separate the background in order to estimate the true  $u_0(x)$ . In order to reconstruct this background and extract the true object  $u_0$  in our work, we perform an additional calibration procedure.

At first glance, this problem looks trivial: one may produce the experiments with a known invariant  $u_0(x)$ , for instance  $u_0(x) = 1$ , obtain the estimate  $\hat{u}_B(x)$  and then recalculate the estimate for the true object as  $\hat{u}_0(x) = \tilde{u}_0(x)/\hat{u}_B(x)$ . However, *a priori* information about the object (its binary amplitude) in the used sparse modeling concerns the true object wave field  $u_0$  but not the disturbed one  $\tilde{u}_0$ . Thus, we are processing the recalculated object estimate at each iteration ( $\hat{u}_0^t(x) = \tilde{u}_0^t(x)/\hat{u}_B(x)$ ,  $t = 0, 1, 2, \dots$ ), and the structure of the developed iterative phase-retrieval algorithm is therefore essentially different from the trivial guess.

The paper is organized as follows. In Section 2, the image formation in a  $4f$  optical system, the model of the object to be reconstructed, and the observation model are presented. The constrained variational approach for the phase retrieval and the sparse modeling for the object phase and amplitude are introduced in Section 3. The proposed phase-retrieval algorithm with the background compensation especially for a binary amplitude object is presented in Section 4. Numerical experiments for

the object wave field reconstructions from real data are shown and discussed in detail in Section 5.

## 2. Observation Model

The straightforward experimental optical setup for the free-space diffraction propagation to different distances  $\{z_r\}$  can be realized by a moveable CCD sensor (see [13, Fig. 1] or [24, Fig. 1]). In such an arrangement, a digital camera is sequentially moved between  $K$  measurement planes (separated by a distance  $\Delta_z$ ) using a motorized precision stage [12,13]. Despite the simplicity of the mathematical model of the free-space propagation, in practice such an optical system is bulky and expensive due to the use of a motorized sensor, and the recording process of intensity measurements is relatively slow. It is shown in [15,25] that a  $4f$  configuration can be used to imitate the lensless optical system for the multi-plane phase-retrieval scenario.

Let us consider the image-formation model in a conventional  $4f$  configuration of the coherent imaging system linking complex amplitudes at the object and measurement planes. The used  $4f$  optical system is illustrated in Fig. 1. Let us denote complex amplitudes at the object and measurement (sensor) planes by  $u_0(x)$  and  $u_r(y)$ , respectively. The lenses  $L_1$  and  $L_2$  with the focal length  $f$  arranged in the  $4f$  configuration provide an accurate mapping of the object wave field into the parallel measurement plane. A reflective phase-modulating SLM is placed at the Fourier plane of the first lens [15]. The forward wave field propagation is realized by the modulation of the Fourier transform of the object wave field using this reprogrammable SLM. The principal feature of this setup is that the sensor plane is immobile and fixed at the distance  $4f$  from the object plane.

Let us assume for a moment that there are no distortions in the optical track. It is well known that the link between the wave fields at the object  $u_0(x)$  and the Fourier planes  $u_F(v/\lambda f)$  is given as follows [14]:

$$u_F\left(\frac{v}{\lambda f}\right) = \frac{1}{i\lambda f} \mathcal{F}\{u_0(x)\}\left(\frac{v}{\lambda f}\right), \quad (2)$$

where  $\mathcal{F}\{\cdot\}$  stands for the two-dimensional (2D) integral Fourier transform, and  $\lambda$  is a wavelength. If the optical mask (SLM) placed at the Fourier plane has the complex-valued transmittance  $M_r(v/\lambda f)$ , then the output of the optical system is defined as

$$u_r(y) = \frac{1}{i\lambda f} \mathcal{F}\left\{u_F\left(\frac{v}{\lambda f}\right) \cdot M_r\left(\frac{v}{\lambda f}\right)\right\}(-y). \quad (3)$$

All these wave field distributions are given in the 2D lateral coordinates: here we use the variables  $x, y, v \in \mathbb{R}^2$  for the object, sensor, and Fourier planes, respectively.

## A. Discrete Modeling

For discrete modeling, the continuous arguments are changed by the digital ones with a corresponding replacement of the continuous functions by their discrete counterparts:  $u_0(x) \rightarrow u_0(k_1\Delta_{x_1}, k_2\Delta_{x_2})$ ,  $u_r(y) \rightarrow u_r(l_1\Delta_{y_1}, l_2\Delta_{y_2})$ ,  $u_F(v/\lambda f) \rightarrow u_F((\Delta_{v_1}/\lambda f)\eta_1, (\Delta_{v_2}/\lambda f)\eta_2)$  with 2D integer arguments  $k = (k_1, k_2)$ ,  $l = (l_1, l_2)$  and  $\eta = (\eta_1, \eta_2)$ . This discretization is dictated by the use of a digital camera and a pixelated SLM as a 2D array of liquid-crystal cells. We hereafter consider the discrete wave fields at the object  $u_0[k_1, k_2]$ , Fourier  $u_F[\eta_1, \eta_2]$ , and sensor  $u_r[l_1, l_2]$  planes.

We use a vector-matrix notation for complex-valued distributions of the wave fields. 2D discrete distributions (matrices) are vectorized to the complex-valued column vector [26]. Bold lowercase characters are used for the vectors. Matrices are defined by bold uppercase characters to distinguish them from vectors. Thus,  $\mathbf{u}_0[k]$ ,  $\mathbf{u}_F[\eta]$  and  $\mathbf{u}_r[l]$  are column vectors constructed by vectorization of the corresponding 2D discrete wave field distributions at the object  $\mathbf{U}_0[k_1, k_2]$ , Fourier  $\mathbf{U}_F[\eta_1, \eta_2]$ , and sensor  $\mathbf{U}_r[l_1, l_2]$  planes, respectively.

In general, these 2D discrete wave field distributions (images)  $\mathbf{U}_0[k_1, k_2]$ ,  $\mathbf{U}_F[\eta_1, \eta_2]$ , and  $\mathbf{U}_r[l_1, l_2]$  can have various pixel sizes  $\Delta_{x_1} \times \Delta_{x_2}$ ,  $\Delta_{v_1} \times \Delta_{v_2}$ , and  $\Delta_{y_1} \times \Delta_{y_2}$ , respectively. In this work we assume that the pixel size at the object and sensor planes is the same:  $\Delta_{x_1} = \Delta_{y_1}$ ,  $\Delta_{x_2} = \Delta_{y_2}$ . Moreover these images can be rectangular of different size  $N_{x_1} \times N_{x_2}$ ,  $N_{v_1} \times N_{v_2}$ , and  $N_{y_1} \times N_{y_2}$ , respectively. Here they are assumed to be of the same size  $N_x \times N_y$  for all planes.

Let us also assume that the following standard conditions are fulfilled [23,27]:

$$\Delta_{v_1} \Delta_{x_1} N_x = \lambda f, \quad \Delta_{v_2} \Delta_{x_2} N_y = \lambda f. \quad (4)$$

Then, discretization of Eq. (3) has the form

$$\mathbf{U}_r[l_1, l_2] = -\frac{1}{N_x N_y} \mathbf{F}\{\mathbf{F}\{\mathbf{U}_0\} \circ \mathbf{M}_r\}[-l_1, -l_2], \quad (5)$$

where  $\mathbf{F}\{\cdot\}$  denotes the 2D discrete Fourier transform,  $\circ$  stands for the Hadamard (elementwise) product, and  $\mathbf{M}_r[\eta_1, \eta_2]$  is the discretized optical mask at the Fourier plane. Note that the fast Fourier transform enables much faster computation of this wave field propagation (see [23,24]).

Taking into account the distortions in the real optical system, the forward wave field propagation from the object to the sensor plane is represented in the form

$$\mathbf{u}_r = \mathbf{A}_r \cdot \tilde{\mathbf{u}}_0, \quad r = 1, \dots, K, \quad (6)$$

where according to Eq. (1),  $\tilde{\mathbf{u}}_0 = \mathbf{u}_0 \circ \mathbf{u}_B$  is a complex-valued vector, corresponding to the disturbed object discrete 2D wave field distribution, and  $\mathbf{u}_B \in \mathbb{C}^{n \times 1}$  is the vector of cumulative disturbances of the optical track (background) recalculated with respect to the

object plane.  $\mathbf{A}_r \in \mathbb{C}^{n \times n}$  is a forward propagation operator corresponding to the optical mask  $\mathbf{M}_r$  at the Fourier plane,  $n = N_x \cdot N_y$ , and  $K$  is a number of these various optical masks.

In general, these could be arbitrary  $\{\mathbf{M}_r\}$ . Following [15], the effect of the object diffraction propagation to various distances  $z_r$  is, however, obtained using the SLM, where the corresponding programmed optical masks  $\mathbf{M}_r$  are

$$\mathbf{M}_r[\eta_1, \eta_2] = \exp \left( i \frac{2\pi}{\lambda} z_r \sqrt{1 - \Delta_{v_1}^2 \frac{|\eta_1|^2}{f^2} - \Delta_{v_2}^2 \frac{|\eta_2|^2}{f^2}} \right). \quad (7)$$

### B. Object Model

In this work we consider the binary amplitude of the object  $\mathbf{u}_0$  defined as follows:

$$\mathbf{a}_0[k] = \text{abs}(\mathbf{u}_0[k]) = \begin{cases} \beta_1, & \text{for } k \in X_1 \subset X, \\ \beta_0, & \text{for } k \in X_0 \subset X, \end{cases} \quad (8)$$

where  $X$  is a support of the image.  $\beta_0 \in \mathbb{R}^+$  and  $\beta_1 \in \mathbb{R}^+$  stand for the nonnegative real-valued lower and upper levels of the object amplitude, respectively. The set  $X_1$  defines the indices of the upper level, and the set  $X_0 = X \setminus X_1$  defines the indices of the lower level. Both the levels  $\beta_0, \beta_1$  and the sets  $X_0, X_1$  are unknown and should be reconstructed. In our experiments we use the U.S. Air Force resolution test chart, which has a binary amplitude. The laser beam passing through the transparent text chart undergoes some phase transformations that reflect in the phase object properties. Thus, we reconstruct the complex-valued object: the binary amplitude [according to Eq. (8)] and an unknown phase  $\boldsymbol{\varphi}_0 = \text{angle}(\mathbf{u}_0)$ .

### C. Noisy Intensity Observations

Assume that we have a set of  $K$  experiments produced with different masks  $\{\mathbf{M}_r\}_r, r = 1, \dots, K$ . The problem is to reconstruct the complex-valued object wave field  $\mathbf{u}_0$  from multiple noisy intensity observations  $\{\mathbf{o}_r\}$  measured at the sensor plane. These measurements are represented in the vector-matrix notation as follows:

$$\mathbf{o}_r[l] = |\mathbf{u}_r[l]|^2 + \varepsilon_r[l], \quad r = 1, \dots, K, \quad (9)$$

where the noise is assumed to be zero-mean Gaussian with the variance  $\sigma_r^2$ ,  $\varepsilon_r[l] \sim \mathcal{N}(0, \sigma_r^2)$ , independent for different  $l$  and  $r$ . The observation vectors  $\{\mathbf{o}_r\}$  correspond to the 2D intensity distributions on the regular discrete grid of the sensor located at the sensor plane.

## 3. Sparse Object Modeling and Variational Formulation

It is assumed in a sparse modeling approach that the *true* object distribution  $\mathbf{u}_0$  can be approximated by a

small number of nonzero elements of basis functions. The ideal basis functions for the object approximation are unknown *a priori* and selected from a given set of potential bases (dictionaries).

### A. Sparse Modeling of Complex-Valued Object

Since we deal with a complex-valued object wave field of the form  $\mathbf{u}_0 = \mathbf{a}_0 \circ \exp(i \cdot \boldsymbol{\varphi}_0)$ , the sparse approximations are performed separately for both the object phase  $\boldsymbol{\varphi}_0 = \text{angle}(\mathbf{u}_0) \in \mathbb{R}^n$  and amplitude  $\mathbf{a}_0 = \text{abs}(\mathbf{u}_0) \in \mathbb{R}^n$  [22–24]. Sparse object approximation can be given in the analysis or synthesis form as follows:

$$\boldsymbol{\theta}_\alpha = \Phi_\alpha \cdot \mathbf{a}_0, \quad \boldsymbol{\theta}_\varphi = \Phi_\varphi \cdot \boldsymbol{\varphi}_0 \quad (\text{analysis}), \quad (10)$$

$$\mathbf{a}_0 = \Psi_\alpha \cdot \boldsymbol{\theta}_\alpha, \quad \boldsymbol{\varphi}_0 = \Psi_\varphi \cdot \boldsymbol{\theta}_\varphi \quad (\text{synthesis}). \quad (11)$$

Here  $\Psi_\alpha, \Psi_\varphi$  and  $\Phi_\alpha, \Phi_\varphi$  are the frame transform matrices, which contain the values of the basis functions for the synthesis and analysis of the object amplitude and phase. The vector  $\boldsymbol{\theta}_\alpha, \boldsymbol{\theta}_\varphi \in \mathbb{R}^m$  can be considered as a spectrum ( $m \gg n$ ) in a parametric data adaptive approximation. Subindices  $\alpha$  and  $\varphi$  are shown for the amplitude and phase, respectively. In particular, it can be said that the phase  $\boldsymbol{\varphi}_0$  and amplitude  $\mathbf{a}_0$  are approximated by small numbers of nonzero components of  $\boldsymbol{\theta}_\varphi$  in the basis  $\Psi_\varphi$  and  $\boldsymbol{\theta}_\alpha$  in the basis  $\Psi_\alpha$ , respectively.

The sparsity of approximation is characterized by either the  $\ell_0$  norm  $\|\boldsymbol{\theta}\|_0$  defined as a number of nonzero components of the vector  $\boldsymbol{\theta}$  or the  $\ell_1$  norm as a sum of absolute values of components of the vector  $\|\boldsymbol{\theta}\|_1 = \sum_s |\theta_s|$ . A smaller value of the norm means a higher sparsity of approximation. The main intention is to find sparsest (shortest) models for phase and amplitude with the smallest values of the  $\ell_0$  or  $\ell_1$  norms [28,29].

Note that results obtained by  $\ell_0$  or  $\ell_1$  norms are shown to be close to each other [30]. For simplicity we use the convex  $\ell_1$  norm in our modeling. The separate sparse modeling for the object phase and amplitude is realized via the powerful block-matching three-dimensional (BM3D)-frame filter, specified for denoising and other imaging problems. We refer to [31–33] for details about the used basis functions, the structure of the frame transform matrices, and the BM3D filter.

### B. Multiobjective Optimization

In our previous works [22–24] the phase-retrieval algorithms are derived from the multiobjective optimization. It is shown in [33] that such an approach is much more efficient and convenient than the optimization of a single constrained criterion, which specifies a number of different subjects/conditions (e.g., the forward diffraction propagation model and the object sparse approximation). We decouple the inverse reconstruction of the disturbed object and filtering of the object amplitude and phase. This

results in simpler implementation. Assume that the background wave field  $\mathbf{u}_B$  is given. Taking into account Eq. (1) and using the separate sparse regularization of the object amplitude and phase, the true object is found by the alternating minimization of two criterion functions:  $\mathcal{J}_1$  with respect to the disturbed object and  $\mathcal{J}_2 = \mathcal{J}_{2,a} + \mathcal{J}_{2,\varphi}$  regarding to the spectra of the true object. The criterion  $\mathcal{J}_1$  is given in the form

$$\begin{aligned} \mathcal{J}_1(\{\mathbf{o}_r\}, \tilde{\mathbf{u}}_0, \{\mathbf{u}_r\}, \{\Lambda_r\}, \tilde{\mathbf{v}}_0) \\ = \sum_{r=1}^K \frac{1}{\sigma_r^2} \left[ \frac{1}{2} \|\mathbf{o}_r - |\mathbf{u}_r|^2\|_2^2 + \frac{1}{\gamma_r} \|\mathbf{u}_r - \mathbf{A}_r \cdot \tilde{\mathbf{u}}_0\|_2^2 \right. \\ \left. + \frac{2}{\gamma_r} \operatorname{Re}\{\Lambda_r^H \cdot (\mathbf{u}_r - \mathbf{A}_r \cdot \tilde{\mathbf{u}}_0)\} \right] + \frac{1}{\gamma_0} \|\tilde{\mathbf{u}}_0 - \tilde{\mathbf{v}}_0\|_2^2. \quad (12) \end{aligned}$$

The first summand in  $\mathcal{J}_1$  is the quadratic fidelity term corresponding to the observation model [Eq. (9)] with the zero-mean Gaussian noise. The following quadratic and linear penalties correspond to the forward diffraction propagation (6), and they are involved with the same positive parameters  $1/\gamma_r$ . Here  $(\cdot)^H$  stands for the Hermitian conjugate,  $\{\Lambda_r\} \in \mathbb{C}^n$  are the complex-valued vectors of the Lagrange multipliers (see [34]), and  $\mathcal{J}_1$  becomes therefore the augmented Lagrangian objective function with respect to the constraint concerning the diffraction propagation model.  $\mathbf{v}_0 = \Psi_a \boldsymbol{\theta}_a \circ \exp(i \cdot \Psi_\varphi \boldsymbol{\theta}_\varphi)$  is an approximation of the complex-valued object  $\mathbf{u}_0$ .  $\tilde{\mathbf{u}}_0 = \mathbf{u}_0 \circ \mathbf{u}_B$ ,  $\tilde{\mathbf{v}}_0 = \mathbf{v}_0 \circ \mathbf{u}_B$ , and the last quadratic penalty corresponding to the synthesis (11) is involved with the parameters  $1/\gamma_0$ .

The criteria  $\mathcal{J}_{2,a}$  and  $\mathcal{J}_{2,\varphi}$  are defined as follows:

$$\mathcal{J}_{2,a}(\boldsymbol{\theta}_a, \mathbf{a}_0) = \frac{1}{2} \|\boldsymbol{\theta}_a - \Phi_a \cdot \mathbf{a}_0\|_2^2 + \tau_a \cdot \|\boldsymbol{\theta}_a\|_1, \quad (13)$$

$$\text{s.t. } \mathbf{a}_0[k] = \begin{cases} \beta_1, & \text{for } k \in X_1, \\ \beta_1, & \text{for } k \in X_0, \end{cases} \quad (14)$$

$$\mathcal{J}_{2,\varphi}(\boldsymbol{\theta}_\varphi, \boldsymbol{\varphi}_0) = \frac{1}{2} \|\boldsymbol{\theta}_\varphi - \Phi_\varphi \cdot \boldsymbol{\varphi}_0\|_2^2 + \tau_\varphi \cdot \|\boldsymbol{\theta}_\varphi\|_1. \quad (15)$$

The first terms in Eqs. (13) and (15) corresponds to the analysis Eqs. (10), and they are calculated for the true object amplitude and phase. The positive parameters  $\tau_a$  and  $\tau_\varphi$  define a balance between the fit of the calculated spectra of the object amplitude and phase and the complexity of the used model. The additional constraint (14) is difficult to be rearranged by a penalty; thus the computation of the true object amplitude is realized with a spacial algorithm.

#### 4. Proposed Algorithm

The flowchart of the proposed phase-retrieval algorithm is shown in Fig. 2. Here we estimate the background using the free-space object and then

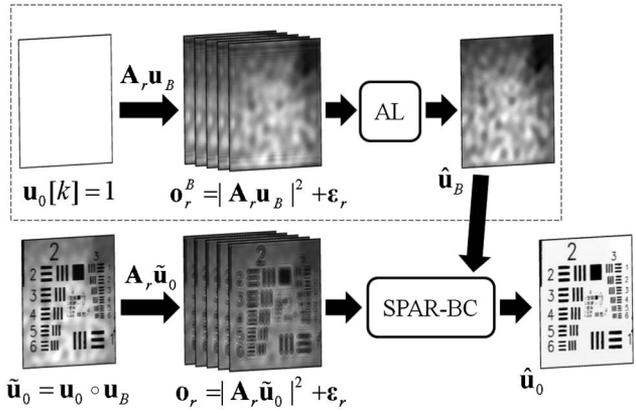


Fig. 2. Flowchart of the proposed phase-retrieval technique with the background compensation. The upper block highlighted by a dashed line represents the background calibration procedure, where the complex-valued estimate of  $\mathbf{u}_B$  is found by AL [34]. The reconstruction of the object using the background estimate is obtained by the proposed SPAR-BC algorithm.

extract the true object. It is assumed that two sets of experiments are made consistently under the same condition: the calibration procedure giving  $\{\mathbf{o}_r^B\}$  to find  $\hat{\mathbf{u}}_B$  and recording of  $\{\mathbf{o}_r\}$  used to compute  $\hat{\mathbf{u}}_0$ .

##### A. Background Reconstruction

We record a number of intensity observations  $\{\mathbf{o}_r^B\}$  with various masks  $\{\mathbf{M}_r\}$  for the free-space object (test image  $\mathbf{u}_0[k] = 1$ ), and find a complex-valued estimate  $\hat{\mathbf{u}}_B$  ( $\tilde{\mathbf{u}}_0 \equiv \mathbf{u}_B$ ) by optimizing the criterion function

$$\begin{aligned} \mathcal{J}_{AL} = \sum_{r=1}^K \frac{1}{\sigma_r^2} \left[ \frac{1}{2} \|\mathbf{o}_r^B - |\mathbf{u}_r|^2\|_2^2 + \frac{1}{\gamma_r} \|\mathbf{u}_r - \mathbf{A}_r \cdot \mathbf{u}_B\|_2^2 \right. \\ \left. + \frac{2}{\gamma_r} \operatorname{Re}\{\Lambda_r^H \cdot (\mathbf{u}_r - \mathbf{A}_r \cdot \mathbf{u}_B)\} \right] + \mu \cdot \|\mathbf{u}_B\|_2^2. \quad (16) \end{aligned}$$

The estimate  $\hat{\mathbf{u}}_B$  can be computed using the augmented Lagrangian (AL) algorithm well described in [34]. This stage is shown in the upper block of Fig. 2 highlighted by a dashed line.

The main difference of  $\mathcal{J}_{AL}$  from  $\mathcal{J}_1$  consists of the last quadratic penalty term.  $\mu$  in Eq. (16) is the Tikhonov regularization parameter that defines a balance between the prior information on  $\mathbf{u}_B$  and the fitting of calculated intensities  $|\mathbf{u}_r|^2$  to the given observations  $\mathbf{o}_r^B$ . Note that in contrast to [34] the criterion  $\mathcal{J}_{AL}$  is presented with respect to the background  $\mathbf{u}_B$  but not to the true object.

##### B. Object Wave Field Reconstruction for Arbitrary Complex-Valued Object

We record intensity measurements for an object  $\{\mathbf{o}_r\}$  using the same masks  $\{\mathbf{M}_r\}$  as before, and reconstruct the true object wave field using the found background estimate  $\hat{\mathbf{u}}_B$ . According to the general idea of the multiobjective optimization, where the alternating minimization of two criteria  $\mathcal{J}_1$  with

respect to  $\tilde{\mathbf{u}}_0 = \mathbf{u}_0 \circ \hat{\mathbf{u}}_B$  and  $\{\mathbf{u}_r\}$  and  $\mathcal{J}_2$  on  $\theta_\alpha$  and  $\theta_\varphi$  are exploited [23,32,33], we use the following iterative algorithm:

$$\mathbf{u}_0^t[k] = \tilde{\mathbf{u}}_0^t[k]/\hat{\mathbf{u}}_B[k], \quad (17)$$

$$\theta_\alpha^t = \arg \min_{\theta_\alpha} \mathcal{J}_{2,\alpha}(\theta_\alpha, \text{abs}(\mathbf{u}_0^t)), \quad (18)$$

$$\theta_\varphi^t = \arg \min_{\theta_\varphi} \mathcal{J}_{2,\varphi}(\theta_\varphi, \text{angle}(\mathbf{u}_0^t)), \quad (19)$$

$$\tilde{\mathbf{v}}_0^t = \Psi_\alpha \theta_\alpha^t \circ \exp(i \cdot \Psi_\varphi \theta_\varphi^t) \circ \hat{\mathbf{u}}_B, \quad (20)$$

$$\mathbf{u}_r^t = \arg \min_{\mathbf{u}_r} \mathcal{J}_1(\mathbf{o}_r, \tilde{\mathbf{u}}_0^t, \mathbf{u}_r, \Lambda_r^t, \tilde{\mathbf{v}}_0^t), \quad (21)$$

$$\Lambda_r^{t+1} = \Lambda_r^t + \alpha_r \cdot (\mathbf{u}_r^t - \mathbf{A}_r \cdot \tilde{\mathbf{u}}_0^t), \quad r = 1, \dots, K, \quad (22)$$

$$\tilde{\mathbf{u}}_0^{t+1} = \arg \min_{\tilde{\mathbf{u}}_0} \mathcal{J}_1(\{\mathbf{o}_r\}, \tilde{\mathbf{u}}_0, \{\mathbf{u}_r^t\}, \{\Lambda_r^t\}, \tilde{\mathbf{v}}_0^t). \quad (23)$$

Here we first estimate the true object [Eq. (17)] by compensation of the disturbed  $\tilde{\mathbf{u}}_0^t$  with the background estimate  $\hat{\mathbf{u}}_B$ . This results in the object amplitude and phase estimates. Then, Eqs. (18) and (19) enable the spectrum estimates of the object amplitude [provided Eq. (8)] and phase by thresholding in the BM3D-frame domain with the thresholds  $\tau_\alpha$  and  $\tau_\varphi$ , respectively [31,32]. Equation (20) corresponds to the synthesis of the approximation  $\tilde{\mathbf{v}}_0^t$  of the disturbed object from the calculated spectra for the true object amplitude and phase and using the background  $\hat{\mathbf{u}}_B$ . Equations (21) and (23) are the minimization steps for  $\mathcal{J}_1$ : the computation of the complex-valued wave field estimates  $\{\mathbf{u}_r\}$  at the sensor planes and the disturbed object  $\tilde{\mathbf{u}}_0$  from noisy intensity observations  $\{\mathbf{o}_r\}$ . The update of the Lagrange variables (derived from the maximization of  $\mathcal{J}_1$  on  $\Lambda_r$ ) is shown in Eq. (22).

This second stage is illustrated in Fig. 2 under the mentioned block for the background estimation.

Suppose there is no prior information on the object amplitude, and hence no additional constraint in the criterion  $\mathcal{J}_{2,\alpha}$ . Then, together the operations in Eqs. (18)–(20) related to the optimization of  $\mathcal{J}_2$  can be rewritten in a more compact form as follows:

$$\begin{aligned} \mathbf{a}_0^{t+1/2} &= \text{BM3D}_\alpha(\text{abs}(\mathbf{u}_0^t)), \\ \boldsymbol{\varphi}_0^{t+1/2} &= \text{BM3D}_\varphi(\text{angle}(\mathbf{u}_0^t)), \\ \tilde{\mathbf{v}}_0^t &= \mathbf{a}_0^{t+1/2} \circ \exp(i \cdot \boldsymbol{\varphi}_0^{t+1/2}) \circ \hat{\mathbf{u}}_B, \end{aligned} \quad (24)$$

where  $\text{BM3D}(\cdot)$  denotes hereafter the processing by the BM3D filter, and the corresponding subindices  $\alpha$  and  $\varphi$  emphasize that the filtering is performed

with different parameters and different transform matrices  $\Psi_\alpha, \Phi_\alpha$  and  $\Psi_\varphi, \Phi_\varphi$  for the amplitude and phase, respectively. In our implementation of the BM3D filter analysis and synthesis operations, the thresholding and calculation of the frame transform matrices are integrated in a single block.

Thus, the algorithm (17)–(23) represents the D–AL from [24] [or the sparse phase amplitude reconstruction (SPAR) algorithm from [23] except the Lagrange multipliers], but with the additional background compensation procedure.

### C. Object Wave Field Reconstruction for Complex-Valued Object with Binary Amplitude

In order to take into consideration the structure of  $\mathbf{a}_0$  and therefore improve the reconstruction quality, a special modification of the filtering procedure (24) is developed.

First, we estimate the subsets  $X_0^t$  and  $X_1^t$ , corresponding to the lower  $\beta_0$  and upper  $\beta_1$  levels of  $\mathbf{a}_0$  [see Eq. (8)], with the thresholding parameter  $\rho^t$  as follows:

$$\begin{aligned} X_0^t &= \{\mathbf{a}_0^t[k] : 0 \leq \mathbf{a}[k]_0^t \leq \rho^t\}, \\ X_1^t &= \{\mathbf{a}_0^t[k] : \mathbf{a}_0^t[k] > \rho^t\}. \end{aligned} \quad (25)$$

This thresholding parameter  $\rho^t$  is calculated using the Otsu algorithm [35]. Second, the estimates  $\beta_1^t$  and  $\beta_0^t$  are computed as medians of  $\mathbf{a}_0^t = \text{abs}(\mathbf{u}_0^t)$  over the found subsets  $X_0^t$  and  $X_1^t$  as

$$\begin{aligned} \beta_0^t &= \text{median}_{\mathbf{a}_0^t[k] \in X_0^t}(\mathbf{a}_0^t[k]), \\ \beta_1^t &= \text{median}_{\mathbf{a}_0^t[k] \in X_1^t}(\mathbf{a}_0^t[k]). \end{aligned} \quad (26)$$

The BM3D filtering of the object amplitude is replaced by

$$\begin{aligned} \mathbf{a}_0^{t+1/3} &= \text{BM3D}_\alpha(\text{abs}(\mathbf{u}_0^t) - \beta_0^t) + \beta_0^t, \\ \mathbf{a}_0^{t+1/2} &= \text{BM3D}_\alpha(\mathbf{a}_0^{t+1/3} - \beta_1^t) + \beta_1^t, \end{aligned} \quad (27)$$

where successive subtractions of  $\beta_0^t$  and  $\beta_1^t$  make the image smoother first in the area of low values of binary object amplitude and after that in the area of its high values.

For the phase filtering we make no partitioning in two subsets as for the object amplitude, because for the considered  $\mathbf{u}_0$  the phase is assumed to be constant, in particular equal to zero. Thus, the only modification [and the difference from Eq. (24)] concerns in the update of the object amplitude. The background compensation and the above-mentioned modification of the BM3D filtering results in the proposed phase-retrieval algorithm especially for an object with a binary amplitude, which is different from D–AL [24] (SPAR [23]).

#### D. Sparse Phase Amplitude Reconstruction with Background Compensation (SPAR-BC)

Assume that the background estimate  $\hat{\mathbf{u}}_B$  is found by AL; then Eqs. (17)–(23) give the advanced phase-retrieval approach with background compensation. Taking into account Eqs. (25)–(27), the reconstruction of the true object wave field is performed by the proposed iterative algorithm defined in the following form:

**Algorithm:** SPAR-BC

**Input:**  $\{\mathbf{o}_r^B\}_{r=1}^K, \{\mathbf{o}_r^I\}_{r=1}^K$

**Initialization:**  $\hat{\mathbf{u}}_B, \tilde{\mathbf{u}}_0^0, \{\Lambda_r^0\}$

**Repeat for**  $t = 0, 1, 2, \dots$

1. Object update (background compensation):

$$\mathbf{u}_0^t[k] = \tilde{\mathbf{u}}_0^t[k]/\hat{\mathbf{u}}_B[k]$$

2. Update of the Otsu's threshold  $\rho^t$

3. Update of  $X_0$  and  $X_1$  [Eq. (25)]

4. Update of the levels  $\beta_0^t$  and  $\beta_1^t$  [Eq. (26)]

5. BM3D filtering:

$$\mathbf{a}_0^{t+1/3} = \text{BM3D}_a(\text{abs}(\mathbf{u}_0^t) - \beta_0^t) + \beta_0^t,$$

$$\mathbf{a}_0^{t+1/2} = \text{BM3D}_a(\mathbf{a}_0^{t+1/3} - \beta_1^t) + \beta_1^t,$$

$$\boldsymbol{\varphi}_0^{t+1/2} = \text{BM3D}_\varphi(\text{angle}(\mathbf{u}_0^t))$$

6. Object approximation synthesis:

$$\mathbf{v}_0^{t+1} = \mathbf{a}_0^{t+1/2} \circ \exp(i \cdot \boldsymbol{\varphi}_0^{t+1/2})$$

**Repeat for**  $r = 1, \dots, K$

7. Forward propagation:

$$\mathbf{u}_r^{t+1/2} = \mathbf{A}_r \cdot \tilde{\mathbf{u}}_0^t$$

8. Observation filtering:

$$\mathbf{u}_r^{t+1}[l] = \mathcal{G}(\mathbf{o}_r[l], \mathbf{u}_r^{t+1/2}[l], \Lambda_r^t[l]) \forall l$$

9. Lagrange multipliers update:

$$\Lambda_r^{t+1} = \Lambda_r^t + \alpha_r \cdot (\mathbf{u}_r^{t+1} - \mathbf{u}_r^{t+1/2})$$

**End on**  $r$

10. Disturbed object update:

$$\tilde{\mathbf{u}}_0^{t+1} = \left( \sum_{r=1}^K \frac{1}{\gamma_r \sigma_r^2} \mathbf{A}_r^H \mathbf{A}_r + \frac{1}{\gamma_0} \mathbf{I} \right)^{-1} \\ \times \sum_{r=1}^K \frac{1}{\gamma_r \sigma_r^2} \mathbf{A}_r^H \cdot (\mathbf{u}_r^{t+1} + \Lambda_r^t) + \frac{1}{\gamma_0} \cdot (\hat{\mathbf{u}}_B \circ \mathbf{v}_0^{t+1})$$

**End on**  $t$

We name this algorithm Sparse Phase Amplitude Reconstruction with Background Compensation (SPAR-BC).

The initialization concerns the calculation of the background (by AL), the initial guess for the disturbed object  $\tilde{\mathbf{u}}_0^0 = \tilde{\mathbf{u}}_0^{\text{init}}$  (e.g., by D-AL [24] or again by AL [34]) and Lagrange multipliers (e.g.,  $\Lambda_r^0[k] = 0$ ). The transform matrices for both the synthesis  $\Psi_a$ ,  $\Psi_\varphi$  and analysis  $\Phi_a$ ,  $\Phi_\varphi$  may be constructed only once during the initialization procedure or may be periodically updated. In this work we calculate these matrices only ones for the compensated object amplitude  $\text{abs}(\tilde{\mathbf{u}}_0^0[k])/\text{abs}(\hat{\mathbf{u}}_B[k])$  and phase estimates  $\text{angle}(\tilde{\mathbf{u}}_0^0[k]) - \text{angle}(\hat{\mathbf{u}}_B[k])$ .

Note that the output of the SPAR-BC phase-retrieval algorithm is not the estimate of the disturbed  $\tilde{\mathbf{u}}_0$  (step 10), but the estimate of the true object wave field  $\mathbf{u}_0$  calculated in step 1. The object amplitude approximation is performed by the consistent application of the Otsu method to separate the values of  $\mathbf{a}_0^t$  into two subsets (steps 2–3), find the estimates of the lower and upper levels (step 4), and filter it with respect to these two estimated levels

(step 5). The derivations of main steps of SPAR-BC [the minimization in Eqs. (18), (19), (21), and (23)] can be found in our previous works [23,24,34]. Step 7 returns the wave field  $\mathbf{u}_r^{t+1/2}$  at the sensor plane corresponding to the forward propagation model with the optical mask  $\mathbf{M}_r$ . Step 8 gives the updates of  $\mathbf{u}_r^{t+1/2}$  by their fitting to the observations  $\mathbf{o}_r$ . The operator defining this update is denoted as  $\mathcal{G}$  and described in [34] [Appendix A].

It is shown in [23] that the object reconstruction with BM3D filtering can be realized without Lagrange multipliers. However, it is found (see [24]) that  $\{\Lambda_r\}$  may essentially help to recover small details of the object. In step 9  $\{\Lambda_r^t\}$  are updated with the step  $\alpha_r$ , and in our experiments we take a fixed step  $\alpha_r = \alpha = 1/20$  for all  $K$  observations.

In this work we use the same noise variation at all sensor planes ( $\sigma_r^2 = \sigma^2$ ) and take the equal parameters for the Lagrangian multipliers,  $\gamma_r = \gamma$ . Then, it is easy to see that the estimate  $\tilde{\mathbf{u}}_0^t$  computed in step 10 of SPAR-BC consists of two parts: the disturbed object estimate calculated from the noisy intensity observations and the filtered object approximation found from the output of the BM3D filter. This step of the algorithm can be given in the form

$$\tilde{\mathbf{u}}_0^{t+1} = \sum_{r=1}^K \mathbf{B}_r \cdot (\mathbf{u}_r^{t+1} + \Lambda_r^t) + \kappa \cdot \tilde{\mathbf{v}}_0^{t+1}, \quad (28)$$

where  $\tilde{\mathbf{v}}_0^{t+1} = \hat{\mathbf{u}}_B \circ \mathbf{v}_0^{t+1}$  and the transform matrix  $\mathbf{B}_r$  is given in the form

$$\mathbf{B}_r = \left( \sum_{r=1}^K \mathbf{A}_r^H \mathbf{A}_r + \kappa \cdot \mathbf{I}_{n \times n} \right)^{-1} \cdot \mathbf{A}_r^H, \quad (29)$$

$\kappa = \gamma \sigma^2 / \gamma_0$ . In particular, for all our experiments  $\kappa = 90/25$ .

#### 5. Numerical Results

In this section an advanced performance of the proposed algorithm is demonstrated by the amplitude reconstruction from noisy disturbed experimental data.

##### A. Settings of Parameters

In our physical experiments one considers  $K = 5$  measurement planes separated from each other by the distance  $\Delta_z = 2$  mm. Thus, in Eq. (7)  $z_r = z_1 + (r - 1) \cdot \Delta_z$  are the distances between the object and sensor planes,  $r = 1, \dots, K$ , and  $z_1 = 20$  mm is the distance from the object to the first measurement plane.

Evidently the experimental results at the sensor plane are different from the output of the model calculated using the angular spectrum decomposition (ASD, [14]) because of the bandlimitedness of the finite-size SLM with the real fill factor of the pixels less than 100%. Despite the fact that the used discrete model of the forward wave field propagation [Eq. (5)] does not accurately reflect the diffraction

propagation in the  $4f$  optical system, the arising difference is considered as a component of the background  $\mathbf{u}_B$  to be estimated and compensated.

In our discrete wave field propagation model, the pixels at the sensor and Fourier planes are square of the different size  $\Delta_{y_1} = \Delta_{y_2} = 3.45 \mu\text{m}$  and  $\Delta_{v_1} = \Delta_{v_2} = 8 \mu\text{m}$ , respectively, with 100% fill factor [36,37]. The object is pixelated with the sensor size pixels:  $\Delta_{x_1} = \Delta_{x_2} = \Delta_{y_1} = \Delta_{y_2}$ . The transparent U.S. Air Force resolution test chart MIL-STD-150A inserted in the front focal plane of the first lens  $L_1$  is illuminated by collimated coherent light with wavelength  $\lambda = 532 \text{ nm}$  (i.e., a green Nd:YAG laser is used). The employed SLM was supplied by Holoeye Photonics AG and configured to provide full  $2\pi$  phase modulation. The focal distance of lenses used in the  $4f$  configuration is  $f = 150 \text{ mm}$ , which equates with the image size  $2892 \times 2892$  pixels according to the sampling conditions (4). The measurement area is smaller and here we reconstruct only a part of the object of the size  $N_x \times N_y$  ( $2048 \times 2048$  pixels) for the corresponding computational focal distance  $f_c = 106.25 \text{ mm}$  [see Eq. (5)]. Note that in Eq. (7), defining the optical masks  $\mathbf{M}_r$ ,  $f = 150 \text{ mm}$ .

The algorithm is implemented for a graphic processing unit (GPU) in order to use the advantage of parallel processing of  $\mathbf{u}_r^{t+1}$  and  $\mathbf{u}_0^{t+1}$ . The GPU realization results in a significant acceleration that is crucial especially for large images [24]. The presented results are computed in MATLAB 7.13 (R2011b) using GPU Nvidia GF460GTX with CUDA 4.1. The computer used for experiments is Intel i5 2500 (four physical cores) at 3.3 GHz, 8Gb RAM, Windows 7 SP1.

#### B. Initialization for SPAR-BC: Reconstruction Without Background Compensation

As mentioned, typically phase-retrieval methods do not incorporate the compensation of the disturbances in the optical system, and in case of real data result in the estimate of  $\tilde{\mathbf{u}}_0$  with imaging that is far from the theoretical prediction. Let us consider the object reconstruction from the experimental data obtained by two fundamentally different multiplane iterative phase-retrieval techniques: by the mentioned parallel AL based algorithm [34] and the successive algorithm described in [38]. The second algorithm is close to the circular wave reconstruction originating in [12,13]: the calculated amplitude is replaced by the square root of the given noisy intensity, keeping the calculated phase (the initial guess for the phase is zero). For simplicity, we refer to the latter algorithm as the Falldorf–Agour (FA) algorithm.

In Fig. 3 we present the comparison of the reconstruction imaging of the disturbed object found by the mentioned methods. In the right column, the estimates of the disturbed object phase and amplitude computed by the FA algorithm [38] are illustrated; see Figs. 3(b) and 3(d), respectively. In the left column, the reconstructed disturbed object phase and amplitude found by AL [34] are demonstrated; see

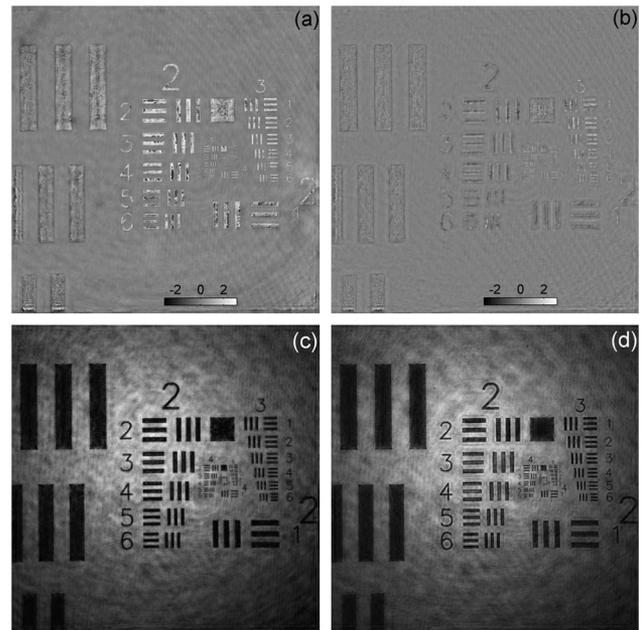


Fig. 3. Reconstructions of the “disturbed” object computed by (left column) AL and (right column) FA [38]. In the top row the amplitude reconstructions are presented by (a) AL and (c) FA. In the bottom row we demonstrate the phase estimates by (c) AL and (d) FA. The AL object reconstruction  $[\tilde{\mathbf{u}}_0^0 = \tilde{\mathbf{a}}_0^0 \cdot \exp(i \cdot \tilde{\varphi}_0^0)]$  with  $\tilde{\mathbf{a}}_0^0$  from (a) and  $\tilde{\varphi}_0^0$  from (c) is used for the initialization of SPAR-BC.

Figs. 3(a) and 3(c), respectively. These quite poor results are obtained using 25 iterations of the algorithms. The artifacts that definitely should be addressed to the background are clearly seen in these images. It can be seen that the amplitude estimate by AL is significantly oversmoothed compared with the result by FA. It is manifested in partial suppression of the diffraction artifacts on the geometrical elements with some degradation of a smooth surface as well. The phase reconstructions are not flat and have certain errors in the regions of the digits and geometrical figures in the amplitude. The phase by AL has stronger degradation of the phase compared with the phase calculated by FA, because of a weak correction of the object reconstruction by Lagrange multipliers (note that the correction of the phase by larger  $\alpha$  leads to more noisy amplitude reconstruction or oversmoothing).

It is recognized that even the powerful BM3D-frame filtering can not extract a sharp object imaging; see, e.g., Figs. 5 and 6 in [24] for the binary amplitude object reconstructions. These imperfect object estimates, in particular the AL reconstruction, are just used as initializations  $\tilde{\mathbf{u}}_0^0 = \tilde{\mathbf{a}}_0^0 \cdot \exp(i \cdot \tilde{\varphi}_0^0)$  for the main procedure of the SPAR-BC algorithm.

#### C. SPAR-BC: Reconstruction with Background Compensation

The background reconstruction is produced with the calibration experiments for the free-space object  $\mathbf{u}_0[k] = 1$ . The reconstructed background amplitude appears quite noisy; thus an additional BM3D post-filtering of the found background  $\hat{\mathbf{u}}_B$  is introduced.

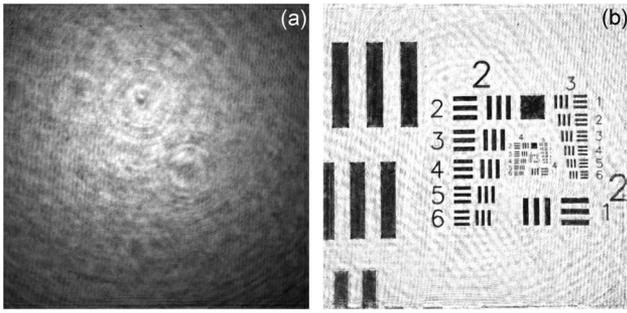


Fig. 4. (a) Amplitude of the smoothed background  $\text{BM3D}_\alpha(\text{abs}(\hat{\mathbf{u}}_B))$ . (b) Initial guess for the object amplitude  $\text{abs}(\hat{\mathbf{u}}_0^0)$  found with the smoothed background amplitude  $\text{abs}(\hat{\mathbf{u}}_0^0)[k]/\text{BM3D}_\alpha(\text{abs}(\hat{\mathbf{u}}_B))[k]$ .

We use the smoothed version of this background reconstruction with no high-frequency components. Figure 4(a) demonstrates the filtered amplitude of the background  $\text{BM3D}_\alpha(\text{abs}(\hat{\mathbf{u}}_B))$ . The cross sections of the original  $\hat{\mathbf{a}}_B = \text{abs}(\hat{\mathbf{u}}_B)$  and smoothed  $\hat{\mathbf{a}}_B = \text{BM3D}_\alpha(\hat{\mathbf{a}}_B)$  background amplitudes are illustrated in Fig. 5. The result of the compensation of the initial object amplitude by such a smoothed background  $\mathbf{a}_0^0[k] = \text{abs}(\hat{\mathbf{u}}_0^0[k])/\hat{\mathbf{a}}_B[k]$  is shown in Fig. 4(b). The corresponding cross section is presented in Fig. 7.

The results of the object reconstruction calculated by the developed SPAR-BC algorithm using 25 iterations are shown in Fig. 6. In Figs. 6(a) and 6(b), the object amplitude reconstruction found with the smoothed and original background estimates, respectively, is presented. The reconstructed object phase angle  $\hat{\mathbf{u}}_0$  is illustrated in Fig. 6(d). The threshold parameter of the BM3D filtering in SPAR-BC is  $\tau_\alpha = 0.13$  for the object amplitude and  $\tau_\varphi = 2$  for the phase.

It can be seen that the imaging of the amplitude estimate obtained using the smoothed background [Fig. 6(a)] is essentially better compared with the reconstruction computed by the background without postfiltering: the artifacts are well seen on the border of Fig. 6(b). Further improvement of the imaging can be achieved by the additional BM3D postfiltering of the reconstructed  $\hat{\mathbf{u}}_0$ : in Fig. 6(c), remaining noise and artifacts are shown to be suppressed. It provides crisp imaging, but the resulting amplitude is over-smoothed and small details are almost lost. The result of postfiltering of  $\hat{\mathbf{a}}_0$  by 12 iterations with

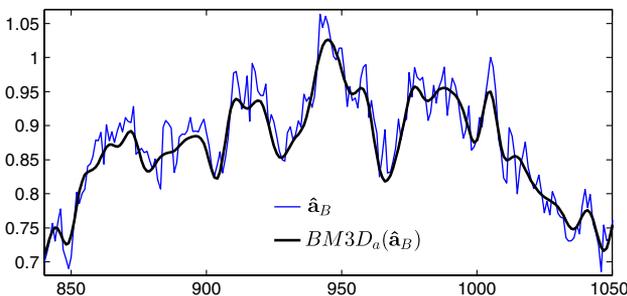


Fig. 5. (Color online) Cross sections of the amplitude estimate of the background [thin curve,  $\hat{\mathbf{a}}_B = \text{abs}(\hat{\mathbf{u}}_B)$ ] and its smoothed version computed by a BM3D filter [thick curve,  $\text{BM3D}_\alpha(\hat{\mathbf{a}}_B)$ ].

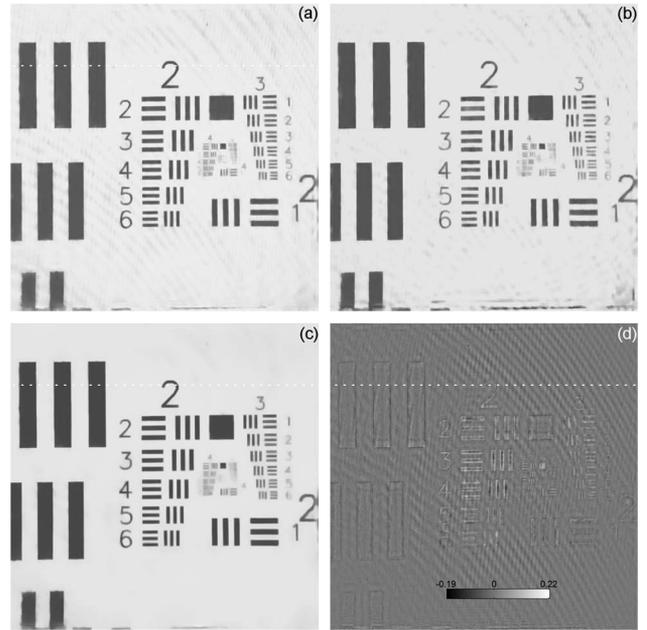


Fig. 6. Object reconstruction by SPAR-BC. The reconstruction of the object amplitude  $\text{abs}(\hat{\mathbf{u}}_0)$  with (a) smoothed background  $\text{BM3D}_\alpha(\text{abs}(\hat{\mathbf{u}}_B))$  and (b) original background  $(\text{abs}(\hat{\mathbf{u}}_B))$ . (c) Result of postfiltering of the object amplitude  $\text{BM3D}_\alpha(\text{abs}(\hat{\mathbf{u}}_0))$  [given in (a)],  $\tau_\alpha = 0.04$ . (d) Object phase estimate angle  $\hat{\mathbf{u}}_0$  with the smoothed background.

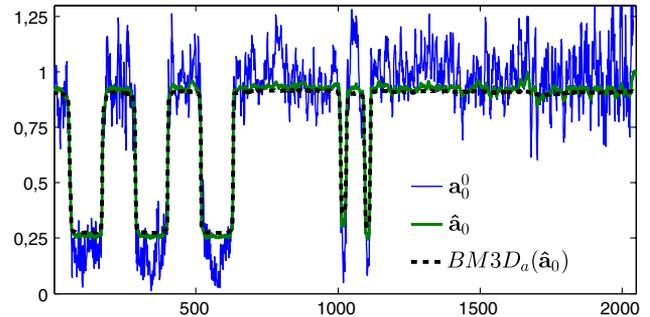


Fig. 7. (Color online) Cross sections of (solid thin curve) the initial guess  $\mathbf{a}_0^0 = \text{abs}(\hat{\mathbf{u}}_0^0)$ , (solid thick) the reconstructed amplitude  $\hat{\mathbf{a}}_0 = \text{abs}(\hat{\mathbf{u}}_0)$ , and (dashed curve) the filtered amplitude estimate  $\text{BM3D}_\alpha(\hat{\mathbf{a}}_0)$ . These results are related to the imaging presented in Figs. 4(b), 6(a), and 6(c), respectively. The cross sections are given along the dashed line in Fig. 6.

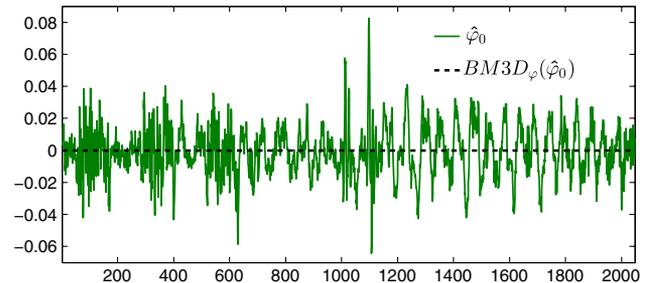


Fig. 8. (Color online) Cross sections of the reconstructed object phase  $\hat{\varphi}_0 = \text{angle}(\hat{\mathbf{u}}_0)$  and the filtered phase  $\text{BM3D}_\varphi(\hat{\varphi}_0)$ ,  $\tau_\varphi = 0.08$ . These results are shown along the dashed lines shown in Fig. 6(d).

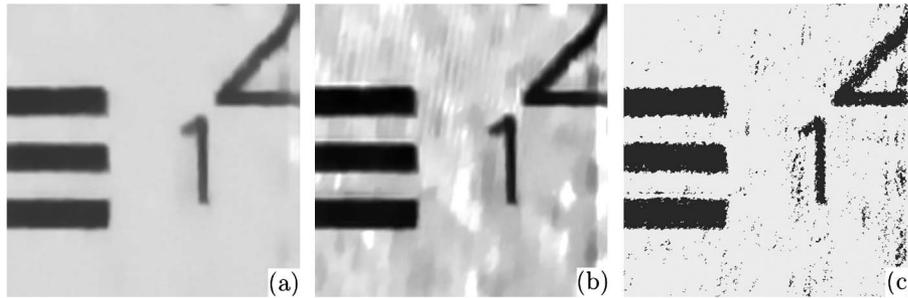


Fig. 9. Fragments ( $384 \times 384$ ) of  $\hat{\mathbf{a}}_0$  with different updates of the object amplitude in SPAR-BC: (a) flattening and BM3D filtering (original SPAR-BC), (b) BM3D filtering only (which is similar to D-AL [24] with the background compensation), and (c) binarization only. The original SPAR-BC (a) enables a relatively flat surface with no significant noise with some oversmoothing. Exclusion of the flattening in BM3D filtering (b) results in a strong degradation of the reconstructing levels. The binarization only used to estimate  $\hat{\mathbf{a}}_0$  (c) leads to more contrast imaging but strong corruption by impulse noise.

$\tau_a = 0.04$  is presented in Fig. 6(c) and the corresponding cross section in Fig. 7.

As for the phase, even with a large thresholding  $\tau_\varphi = 2$  we have relatively large noise in the phase estimate [see Fig. 6(d)], because at each iteration we deal with noisy estimates of  $\mathbf{u}_r$  computed from the given intensity observations [see Eq. (28)]. Assuming that  $\varphi_0[k] = 0$ , RMSE = 0.2 for the object estimate. However, we can completely wipe the phase noise out by (again) the mentioned additional BM3D post-filtering with quite a small  $\tau_\varphi = 0.08$ : compare the cross sections of  $\hat{\varphi}_0$  and  $\text{BM3D}_{\varphi}(\hat{\varphi}_0)$  in Fig. 8.

## 6. Discussion and Conclusion

In general, there are at least two alternatives for the update of the object amplitude in the proposed SPAR-BC algorithm (steps 2–5). If one ignores *a priori* information on the structure of  $\mathbf{a}_0$  and uses Eqs. (24) without the modifications given in Eqs. (25)–(27), then steps 2–4 vanish and SPAR-BC becomes the D-AL algorithm from [24] with the background compensation procedure. On other hand, the BM3D filtering of  $\mathbf{a}_0$  can be replaced by only direct binarization:  $\mathbf{a}_0^{t+1/2}[k] = \beta_0^t$ , if  $\mathbf{a}_0^{t+1/2}[k] \in X_0^t$  and  $\mathbf{a}_0^{t+1/2}[k] = \beta_1^t$ —otherwise. In the latter case BM3D filtering of the object amplitude is found to be not productive, and it vanishes from step 5 of SPAR-BC. In Fig. 9 we illustrate fragments of the amplitude

reconstructions  $\hat{\mathbf{a}}_0$  by these approaches: by SPAR-BC [Fig. 9(a)], by such D-AL with the background compensation with BM3D filtering only [Fig. 9(b)], and by the mentioned binarization only, without filtering of the amplitude [Fig. 9(c)]. The corresponding cross sections are demonstrated in Fig. 10. It can be seen that the upper and lower levels of the reconstructed amplitudes found by the first alternative (D-AL with the background compensation denoted in Fig. 10 by “D-AL with BC”) are far from being flat, which affects the imaging. The result of binarization is corrupted by impulse (salt and pepper) noise, and in spite of more contrast imaging the reconstruction accuracy in terms of RMSE is worse compared with the original SPAR-BC because there are a lot of errors on the upper level, and  $\hat{\mathbf{a}}_0$  jumps between 0.16 and 0.9.

It can be seen that the flattening with the BM3D-frame filtering in SPAR-BC works as a classifier for the noisy binary object estimate. The estimate of the

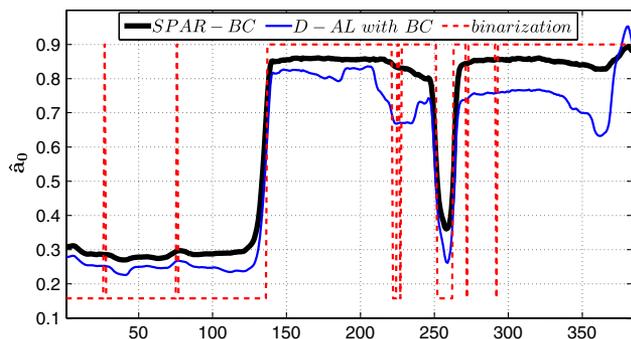


Fig. 10. (Color online) Cross sections of the fragments of  $\hat{\mathbf{a}}_0$  presented in Fig. 9. Here the only use of the BM3D filter for the object amplitude update is denoted by “D-AL with BC.”

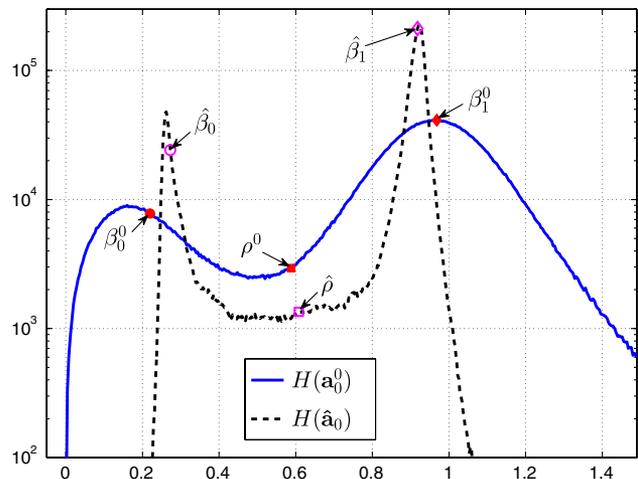


Fig. 11. (Color online) Partition produced according to the flattening with BM3D filtering [Eqs. (25)–(26)].  $H(\mathbf{a}_0^0)$  is the histogram for the initial estimate of the object amplitude (solid curve,  $\mathbf{a}_0^0 = \text{abs}(\mathbf{u}_0^0)$ ), presented in Fig. 4(b).  $H(\hat{\mathbf{a}}_0)$  is the histogram for the resulting object amplitude estimate after 25 iterations [dashed curve,  $\hat{\mathbf{a}}_0 = \text{abs}(\hat{\mathbf{u}}_0)$ ], illustrated in Fig. 6(a). ( $\rho^0$ ,  $\beta_1^0$ ,  $\beta_0^0$ ) and ( $\hat{\rho}$ ,  $\hat{\beta}_1$ ,  $\hat{\beta}_0$ ) are Otsu’s threshold, the upper and lower levels of the initial and resulting object estimates, respectively.

amplitude levels is found using the Otsu method [35], but the BM3D filtering shifts the value of the pixel  $a_0^t[k]$  to one of these two levels  $\beta_0^t$  or  $\beta_1^t$  depending on the local neighborhood. The result of partitioning is presented in Fig. 11. Let  $H(\cdot)$  stand for a histogram of a discrete distribution. The histogram  $H(a_0^0)$  for the initial estimate of the compensated object amplitude is denoted in Fig. 11 by a solid curve, and the histogram of the resulting  $\hat{a}_0$  is denoted by a dashed curve. With 25 iterations we arrive at the obviously binary  $\hat{a}_0$ . In Fig. 11 we also present the estimated levels of the object amplitude (initial  $\beta_0^0, \beta_1^0$  and resulting  $\hat{\beta}_0, \hat{\beta}_1$ ) and the Otsu threshold (for the initialization  $\rho^0$  and the final step  $\hat{\rho}$ ). Note that blurred regions [see small details and borders on the geometrical elements, e.g., in Fig. 6(a)] correspond to a “sloping valley” between two peaks of  $H(\hat{a}_0)$  in Fig. 11.

In this paper, a novel phase-retrieval algorithm with background compensation and powerful BM3D filtering is presented. It is shown that the proposed SPAR-BC algorithm demonstrates a very good reconstruction quality: we have a clear separation of the binary true object, and the background estimate “undertakes” strong fluctuations, which would be difficult to compensate by filtering only [24, cf. Fig. 5]. The reconstructions by two different phase-retrieval methods (AL and FA) are presented to emphasize the obtained enhancement of imaging of the developed algorithm with respect to modern phase-retrieval algorithms with no background compensation (compare the results in Figs. 3 and 6).

This work was supported by the Academy of Finland, project No. 138207, 2011-2014. The post-graduate work of Artem Migukin is funded by the Tampere Doctoral Programme in Information Science and Engineering (TISE). The authors also thank R. B. Bergmann and C. von Kopylow for continuous support and valuable comments. M. Agour gratefully acknowledges the financial support of the Deutsche Forschungsgemeinschaft (DFG) for funding a part of this work within the frame of the project DynaLiFeS (BE1924/2-1).

## References

1. R. W. Gerchberg and W. O. Saxton, “A practical algorithm for the determination of phase from image and diffraction plane pictures,” *Optik* **35**, 237–246 (1972).
2. D. L. Misell, “A method for the solution of the phase problem in electron microscopy,” *J. Phys. D* **6**, L6–L9 (1973).
3. R. A. Gonsalves, “Phase retrieval from modulus data,” *J. Opt. Soc. Am.* **66**, 961–964 (1976).
4. J. R. Fienup, “Reconstruction of an object from the modulus of its Fourier transform,” *Opt. Lett.* **3**, 27–29 (1978).
5. B. Gu and G. Yang, “On the phase retrieval problem in optical and electronic microscopy,” *Acta Opt. Sin.* **1**, 517–522 (1981).
6. G. Yang, B. Dong, B. Gu, J. Zhuang, and O. K. Ersoy, “Gerchberg–Saxton and Yang–Gu algorithms for phase retrieval in a nonunitary transform system: a comparison,” *Appl. Opt.* **33**, 209–218 (1994).
7. Z. Zalevsky, D. Mendlovic, and R. Dorsch, “Gerchberg–Saxton algorithm applied in the fractional Fourier or the Fresnel domain,” *Opt. Lett.* **21**, 842–844 (1996).
8. T. E. Gureyev, “Composite techniques for phase retrieval in the Fresnel region,” *Opt. Commun.* **220**, 49–58 (2003).
9. J. R. Fienup, “Iterative method applied to image reconstruction and to computer generated holograms,” *Opt. Eng.* **19**, 297–305 (1980).
10. J. R. Fienup, “Phase retrieval algorithms: a comparison,” *Appl. Opt.* **21**, 2758–2769 (1982).
11. V. Yu. Ivanov, V. P. Sivokon, and M. A. Vorontsov, “Phase retrieval from a set of intensity measurements: theory and experiment,” *J. Opt. Soc. Am. A* **9**, 1515–1524 (1992).
12. G. Pedrini, W. Osten, and Y. Zhang, “Wave-front reconstruction from a sequence of interferograms recorded at different planes,” *Opt. Lett.* **30**, 833–835 (2005).
13. P. Almero, G. Pedrini, and W. Osten, “Complete wavefront reconstruction using sequential intensity measurements of a volume speckle field,” *Appl. Opt.* **45**, 8596–8605 (2006).
14. J. W. Goodman, *Introduction to Fourier Optics*, 2nd ed. (McGraw-Hill, 1996).
15. C. Falldorf, M. Agour, C. v. Kopylow, and R. B. Bergmann, “Phase retrieval by means of a spatial light modulator in the Fourier domain of an imaging system,” *Appl. Opt.* **49**, 1826–1830 (2010).
16. Q. Xue, Z. Wang, J. Huang, and J. Gao, “The elimination of the errors in the calibration image of 3D measurement with structured light,” *Proc. SPIE* **8430**, 84300N (2012).
17. E. Cuche, P. Marquet, and C. Depeursinge, “Spatial filtering for zero-order and twin-image elimination in digital off-axis holography,” *Appl. Opt.* **39**, 4070–4075 (2000).
18. P. Ferraro, S. D. Nicola, A. Finizio, G. Coppola, S. Grilli, C. Magro, and G. Pierattini, “Compensation of the inherent wave front curvature in digital holographic coherent microscopy for quantitative phase-contrast imaging,” *Appl. Opt.* **42**, 1938–1946 (2003).
19. G. Pedrini, S. Schedin, and H. J. Tiziani, “Aberration compensation in digital holographic reconstruction of microscopic objects,” *J. Mod. Opt.* **48**, 1035–1041 (2001).
20. S. M. Zhao, J. Leach, L. Y. Gong, J. Ding, and B. Y. Zheng, “Aberration corrections for free-space optical communications in atmosphere turbulence using orbital angular momentum states,” *Opt. Express* **20**, 452–461 (2012).
21. S. Grilli, P. Ferraro, S. D. Nicola, A. Finizio, G. Pierattini, and R. Meucci, “Whole optical wavefields reconstruction by digital holography,” *Opt. Express* **9**, 294–302 (2001).
22. A. Migukin, V. Katkovnik, and J. Astola, “Advanced phase retrieval: maximum likelihood technique with sparse regularization of phase and amplitude,” arXiv:1108.3251v1.
23. V. Katkovnik and J. Astola, “Phase retrieval via spatial light modulator phase modulation in  $4f$  optical setup: numerical inverse imaging with sparse regularization for phase and amplitude,” *J. Opt. Soc. Am. A* **29**, 105–116 (2012).
24. A. Migukin, V. Katkovnik, and J. Astola, “Advanced multi-plane phase retrieval using graphic processing unit: augmented Lagrangian technique with sparse regularization,” *Proc. SPIE* **8429**, 84291N (2012).
25. M. Agour, C. Falldorf, C. v. Kopylow, and R. B. Bergmann, “Automated compensation of misalignment in phase retrieval based on a spatial light modulator,” *Appl. Opt.* **50**, 4779–4787 (2011).
26. J. R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, 2nd ed. (Wiley, 1999).
27. Th. Kreis, *Handbook of Holographic Interferometry: Optical and Digital Methods* (Wiley-VCH, 2005).
28. M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing* (Springer, 2010).
29. D. Han, K. Kornelson, D. Larson, and E. Weber, *Frames for Undergraduates* (Student Mathematical Library, AMS, 2007).
30. D. L. Donoho, “Compressed sensing,” *IEEE Trans. Inf. Theory* **52**, 1289–1306 (2006).
31. V. Katkovnik, A. Danielyan, and K. Egiazarian, “Decoupled inverse and denoising for image deblurring: variational BM3D-frame technique,” in *Proceedings of the International Conference on Image Processing (ICIP)* (IEEE, 2011), pp. 3514–3517.
32. A. Danielyan, V. Katkovnik, and K. Egiazarian, “Image deblurring by augmented Lagrangian with BM3D frame

- prior,” presented at Workshop on Information Theoretic Methods in Science and Engineering (WITMSE), Tampere, Finland, 16–18 August 2010.
33. A. Danielyan, V. Katkovnik, and K. Egiazarian, “BM3D frames and variational image deblurring,” *IEEE Trans. Image Process.* **21**, 1715–1728 (2012).
  34. A. Migukin, V. Katkovnik, and J. Astola, “Wave field reconstruction from multiple plane intensity-only data: augmented Lagrangian algorithm,” *J. Opt. Soc. Am. A* **28**, 993–1002 (2011).
  35. N. Otsu, “A threshold selection method from gray-level histograms,” *IEEE Trans. Syst. Man Cybernet.* **9**, 62–66 (1979).
  36. V. Arrizon, E. Carreon, and M. Testorf, “Implementation of Fourier array illuminators using pixelated SLM: efficiency limitations,” *Opt. Commun.* **16**, 207–213 (1999).
  37. M. Agour, C. Falldorf, and C. von Kopylow, “Digital pre-filtering approach to improve optically reconstructed wavefields in opto-electronic holography,” *J. Opt.* **12**, 055401 (2010).
  38. M. Agour, C. Falldorf, C. von Kopylow, and R. B. Bregmenn, “The effect of misalignment in phase retrieval based on a spatial light modulator,” *Proc. SPIE* **8082**, 80820M (2011).