Wave field reconstruction from multiple plane intensity-only data: augmented Lagrangian algorithm

Artem Migukin,* Vladimir Katkovnik, and Jaakko Astola

Department of Signal Processing, Tampere University of Technology, P.O. Box 527, FI-33101 Tampere, Finland
**Corresponding author: artem.migukin@tut.fi*

Received November 5, 2010; revised March 9, 2011; accepted March 9, 2011;
posted March 16, 2011 (Doc. ID 137293); published May 9, 2011

A complex-valued wave field is reconstructed from intensity-only measurements given at multiple observation planes parallel to the object plane. The phase-retrieval algorithm is obtained from the constrained maximum likelihood approach provided that the additive noise is Gaussian. The forward propagation from the object plane to the measurement plane is treated as a constraint in the proposed variational setting of reconstruction. The developed iterative algorithm is based on an augmented Lagrangian technique. An advanced performance of the algorithm is demonstrated by numerical simulations.

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

1. INTRODUCTION

Phase retrieval is a problem of estimating the phase of a complex-valued wave field from intensity observations. The phase is recovered, mainly iteratively, from a number of intensity measurements made at different observation planes. Phase-retrieval algorithms have being exploited in several areas, such as microscopy, crystallography, astronomy, and deformation detection. Experimental arrangements for phase-retrieval methods are often simpler and cheaper than for interferometric ones, which require a reference beam. An important advantage of phase-retrieval techniques is their high robustness to disturbances (e.g., vibration), which degrade the accuracy in interferometry.

In 1972, Gerchberg and Saxton [1] proposed a simple and efficient iterative algorithm for phase retrieval, initially for a single observation plane. Generalization and development of this algorithm for various applications has been done by many authors (e.g., [2,3]). Fienup systematized the earlier works and, in 1982, introduced the following classes of iterative phase-retrieval algorithms: the error-reduction, gradient search, and input–output methods [4].

The mathematical formulation of the phase-retrieval problem, the existence and uniqueness of solution, and the convergence of the algorithms have been comprehensively studied in a number of publications (e.g., [5,6] and references within the papers). New connections between the conventional phase-retrieval algorithms and convex optimization methods were established in [7]. It is shown that these algorithms can be identified as nonconvex alternating projection algorithms. In particular, the iterative Gerchberg–Saxton–Fienup algorithms can be interpreted as the iterative projections at the sensor and object planes. This interpretation highlights the nature of these algorithms as well as gives an opportunity to design novel algorithms with flexible use of an extra information on measurements and reconstructed distributions.

In our work, we consider the problem of the optimal wave field reconstruction from intensity observations. In order to achieve this goal, we use a variational constrained maximum likelihood formulation. The phase-retrieval algorithm is obtained as a solution of this optimization problem.

Let us introduce basic notation and consider the multiplane wave field reconstruction scenario in order to make clear our formulation and review relevant publications.

Assume that $u_0(x)$ and $u_r(x)$, $r = 1, ...K$, denote complex-valued wave field distributions at the object and sensor planes, respectively, given in lateral coordinates $x \in \mathbb{R}^2$. The index r corresponds to a distance $z_r = z_1 + (r-1)\cdot\Delta_z$ between the parallel object and the rth observation plane, where Δ_z is a distance between the observation planes and K is a number of these planes. It is assumed that the wave field distributions at the object and sensor planes are pixelated, i.e., they are pixelwise invariant. In discrete modeling, a continuous variable x is replaced by a digital one with the corresponding replacement of continuous distributions by their discrete counterparts: $u_0(x) \to u_0[k]$, $u_r(x) \to u_r[k]$ with a two-dimensional (2D) integer argument k.

Discrete intensity observations are given in the form

$$o_r[k] = |u_r[k]|^2 + \varepsilon_r[k], \qquad r = 1, ...K, \tag{1}$$

where the wave field intensity (power) is measured with an additive random error $\varepsilon_r[k]$.

For simplicity and referring to the central limit theorem, we assume that the resulting noise ε_r is zero-mean Gaussian, with the standard deviation σ_r for the rth plane, i.e., $\varepsilon_r[k] \sim \mathcal{N}(0, \sigma_r^2)$. This approach can be extended to more complex distributions, such as Poissonian and mixed Poissonian–Gaussian.

The problem at hand is to reconstruct pixelated complex-valued wave field distributions at the object $u_0[k]$ and sensor planes $u_r[k]$ from the noisy intensity data $o_r[k]$. This multiplane setup is illustrated in Fig. 1.

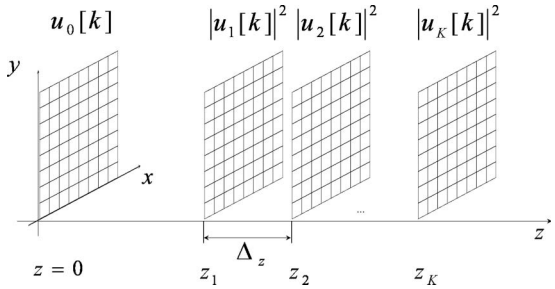

Fig. 1. Multiple plane wave field reconstruction scenario: $u_0[k]$ and $u_r[k]$ are wave field distributions at the object and the rth parallel measurement (sensor) plane, respectively, $r = 1, ...K$.

In this paper, we use a vector-matrix notation for complex-valued distributions of the wave fields at the object and sensor planes as $\mathbb{C}^{n\times 1}$ vectors. For 2D discrete distributions (matrices) of the size $N \times M$, the complex-valued vector variables of the length $n = N \cdot M$ are constructed by concatenating columns of these matrices. Bold lower case characters are used for these vectors.

With this notation, the forward wave field propagation from a diffraction (object) plane with a complex-valued distribution $\mathbf{u}_0$ gives a complex-valued distribution $\mathbf{u}_r$ at the rth image (sensor) plane as

$$\mathbf{u}_r = \mathbf{A}_r \cdot \mathbf{u}_0, \tag{2}$$

where $\mathbf{A}_r$ is the forward propagation operator from the object plane to the rth sensor plane. For $\mathbf{u}_0$ and $\mathbf{u}_r$ of the same sizes, this operator is a complex-valued $n \times n$ matrix, $\mathbf{A}_r \in \mathbb{C}^{n\times n}$.

In the vector-matrix notation, Eq. (1) takes the form

$$\mathbf{o}_r = |\mathbf{u}_r|^2 + \varepsilon_r, \qquad r = 1, ...K, \tag{3}$$

where the modulus $|\cdot|$ and square of modulus $|\cdot|^2$ are the elementwise operations applied to the elements of the corresponding vectors; thus, $|\cdot|$ and $|\cdot|^2$ are vectors composed from the modulus and square modulus of the elements of the corresponding vector.

We consider a coherent light scenario with the paraxial approximation of the wave field propagation based on the Rayleigh–Sommerfield integral [8].

The operator $\mathbf{A}_r$ in Eq. (2) is specified by discretization of this integral. Depending on the discretization used, this discrete forward propagation model can be: convolutional single or double size [9], angular spectrum decomposition (ASD) [8], the recent discrete diffraction transform (DDT) given in the matrix (M-DDT) [10] or frequency domain (F-DDT) forms [11]. These DDT models are obtained for the Fresnel approximation of the Rayleigh–Sommerfield integral and enable an accurate pixel-to-pixel mapping of the pixelated $\mathbf{u}_0$ to $\mathbf{u}_r$.

The wave field $\mathbf{u}_r$ can be generated by the object distribution $\mathbf{u}_0$ [according to Eq. (2)] as well as by the wave field from any previous sensor planes (say, $\mathbf{u}_s$, $z_s < z_r$) [12]:

$$\mathbf{u}_r = \mathbf{A}_{r,s} \cdot \mathbf{u}_s, \tag{4}$$

where $\mathbf{A}_{r,s}$ denotes a propagation operator from the sth to the rth sensor plane.

These two different interpretations of the wave field propagation [Eqs. (2) and (4)] result in the corresponding *parallel* and *successive* algorithms for phase retrieval.

In this paper, the phase-retrieval problem is reduced to the reconstruction of the complex-valued $\mathbf{u}_0$ from the given observations [Eq. (3)]. Then the three-dimensional wave field can be reconstructed for any longitudinal coordinate z, for instance, for any sensor plane.

A. Parallel Algorithms

A multiplane version of the iterative Gerchberg–Saxton–Fienup algorithm can be given in the vector-matrix notation as:

$$\hat{\mathbf{u}}_r^{(p)} = \mathbf{A}_r \cdot \hat{\mathbf{u}}_0^{(p)}, \qquad p = 0, 1, ...,$$
$$\hat{\mathbf{u}}_0^{(p+1)} = \frac{1}{K}\sum_{r=1}^{K} \mathbf{A}_r^H \cdot \left[\frac{\sqrt{\mathbf{o}_r}}{|\hat{\mathbf{u}}_r^{(p)}|} \circ \hat{\mathbf{u}}_r^{(p)}\right], \tag{5}$$

where $\mathbf{A}_r^H = (\mathbf{A}_r^*)^T$ is the Hermitian transpose of $\mathbf{A}_r$, and $\mathbf{o}_r$ are the intensity observations.

In Eq. (5), the operations in the square brackets are Hadamard elementwise, i.e., $\frac{\sqrt{\mathbf{o}_r}}{|\hat{\mathbf{u}}_r^{(p)}|}$ is a vector obtained by the elementwise division of the vector $\sqrt{\mathbf{o}_r}$ by $|\hat{\mathbf{u}}_r^{(p)}|$. Multiplication of the vectors $\frac{\sqrt{\mathbf{o}_r}}{|\hat{\mathbf{u}}_r^{(p)}|}$ by $\hat{\mathbf{u}}_r^{(p)}$ is also elementwise. Thus, $\left[\frac{\sqrt{\mathbf{o}_r}}{|\hat{\mathbf{u}}_r^{(p)}|} \circ \hat{\mathbf{u}}_r^{(p)}\right]$ is a vector. Note that the multiplication $\hat{\mathbf{u}}_r^{(p)}$ by $\frac{\sqrt{\mathbf{o}_r}}{|\hat{\mathbf{u}}_r^{(p)}|}$ means a replacement of the magnitude of $\hat{\mathbf{u}}_r^{(p)}$ by $\sqrt{\mathbf{o}_r}$, keeping the phase of $\hat{\mathbf{u}}_r^{(p)}$.

The algorithm in Eq. (5) works as follows: at the pth iteration, the estimates at the sensor planes are calculated by the object estimate $\hat{\mathbf{u}}_0^{(p)}$, as in Eq. (2). Then the moduli of the estimates $\hat{\mathbf{u}}_r^{(p)}$ are replaced by the observed $\sqrt{\mathbf{o}_r}$ and these updated $\hat{\mathbf{u}}_r^{(p)}$ are propagated backward to the object plane by multiplying these estimates by $\mathbf{A}_r^H$. In this way we obtain K object wave field estimates of the form $\mathbf{A}_r^H \cdot \left[\frac{\sqrt{\mathbf{o}_r}}{|\hat{\mathbf{u}}_r^{(p)}|} \circ \hat{\mathbf{u}}_r^{(p)}\right]$.

The sample mean of these estimates gives the object estimate for the next iteration. These iterations are repeated until convergence.

It is shown for this algorithm that the discrepancy between the measured and reconstructed magnitudes, calculated as $J_1 = \sum_r ||\sqrt{\mathbf{o}_r} - |\mathbf{u}_r|||^2$, can never increase between iterations [4].

A different variational formulation of phase retrieval is developed in [13] with minimization of the divergence between the measured and reconstructed intensities calculated as $J_2 = \sum_r ||\mathbf{o}_r - |\mathbf{u}_r|^2||^2$.

The approach proposed in [14,15] is based on minimization of the discrepancy between measured and reconstructed intensities calculated using the probabilistic relative entropy also known as the Kullback–Leibler ($\mathcal{KL}$) criterion. It assumes that the distributions are random and the criterion is of the form

$$\mathcal{KL} = \sum_{r=1}^{K} \int_{R^2} \mathbf{o}_r(x) \lg \frac{\mathbf{o}_r(x)}{|\mathbf{u}_r(x)|^2} \mathrm{d}x. \tag{6}$$

The estimate is found as a solution of the constrained optimization

$$\hat{\mathbf{u}}_0(x) = \min_{\mathbf{u}_0} \mathcal{KL} \quad \text{provided} \int_{R^2} |\mathbf{o}_r(x)|\mathrm{d}x = \int_{R^2} |\mathbf{u}_r(x)|^2 \mathrm{d}x. \tag{7}$$

The iterative algorithm derived in [14] for Eq. (7) can be represented as

$$\begin{aligned} \hat{\mathbf{u}}_r^{(p)} &= \mathbf{A}_r \cdot \hat{\mathbf{u}}_0^{(p)}, \qquad p = 0, 1, ..., \\ \hat{\mathbf{u}}_0^{(p+1)} &= \frac{1}{K} K \sum_{r=1} \mathbf{A}_r^H \cdot \left[\frac{\mathbf{o}_r}{|\hat{\mathbf{u}}_r^{(p)}|^2} \circ \hat{\mathbf{u}}_r^{(p)} \right], \end{aligned} \tag{8}$$

where operations for $\frac{\mathbf{o}_r}{|\hat{\mathbf{u}}_r^{(p)}|^2} \circ \hat{\mathbf{u}}_r^{(p)}$ are elementwise, as in Eq. (5).

The algorithm in Eq. (8) is different from Eq. (5) by the weights for $\hat{\mathbf{u}}_r^{(p)}$. Instead of the ratio of the moduli in Eq. (5), the ratio of the intensities (squared moduli) appears in Eq. (8).

A discussion on connections between the algorithm in Eq. (8) with the algorithms based on minimization of J_1 and J_2 can be found in [14].

In [16], we presented an algorithm of a structure similar to Eq. (8) but with different propagation operators:

$$\begin{aligned} \hat{\mathbf{u}}_r^{(p)} &= \mathbf{A}_r \cdot \hat{\mathbf{u}}_0^{(p)}, \qquad p = 0, 1, ..., \\ \hat{\mathbf{u}}_0^{(p+1)} &= \left(\sum_{r=1}^{K} \mathbf{A}_r^H \mathbf{A}_r + \mu \cdot \mathbf{I}_{n \times n} \right)^{-1} \\ &\quad \times \sum_{r=1}^{K} \mathbf{A}_r^H \cdot \left[\frac{\sqrt{\mathbf{o}_r}}{|\hat{\mathbf{u}}_r^{(p)}|} \circ \hat{\mathbf{u}}_r^{(p)} \right], \end{aligned} \tag{9}$$

where $\mu > 0$ is a regularization parameter and σ_r^2 is the variance of the noise at the rth plane. This algorithm is derived from the following speculations. The linear least square estimate is obtained for $\hat{\mathbf{u}}_0$, assuming that complex-valued observations are available for the sensor planes. In the algorithm, the moduli of these hypothetically complex-valued data at the sensor planes are replaced by the square roots of the measurements. Note that, for invertible propagation operators (e.g., ASD) $\mathbf{A}_r^H \mathbf{A}_r = \mathbf{I}$, and for $\mu = 0$, the algorithm in Eq. (9) becomes Eq. (5).

The algorithms in Eqs. (5), (8), and (9) belong to the class of the parallel algorithms. A specific feature of this class is that the object distribution is the only unknown variable and all observations are used in parallel for iterative calculation of this variable.

B. Successive Algorithms

Recently, a circular phase-retrieval algorithm is proposed in [17,18]. It is known as the single-beam multiple-intensity phase reconstruction (SBMIR) algorithm. Let the wave fields at the rth and $(r+1)$th planes be linked as $\mathbf{u}_{r+1} = \mathbf{A}_{r+1,r} \cdot \mathbf{u}_r$, where $\mathbf{A}_{r+1,r}$ is an operator (matrix), connecting the wave field distributions at the corresponding planes. Then the forward–forward algorithm as it is defined in [19] can be written in the form

1. Repeat for $p = 0, 1, ...,$
2. For $r = 1, ...K-1,$ $\quad \hat{\mathbf{u}}_{r+1}^{(p)} = \mathbf{A}_{r+1,r} \cdot \left[\frac{\sqrt{\mathbf{o}_r}}{|\hat{\mathbf{u}}_r^{(p)}|} \circ \hat{\mathbf{u}}_r^{(p)} \right],$
3. For $r = K,$ $\quad \hat{\mathbf{u}}_1^{(p+1)} = \mathbf{A}_{1,K} \cdot \left[\frac{\sqrt{\mathbf{o}_K}}{|\hat{\mathbf{u}}_K^{(p)}|} \circ \hat{\mathbf{u}}_K^{(p)} \right],$
4. End on r,
5. End for p. (10)

Here, similar to Eq. (5), the factor $\frac{\sqrt{\mathbf{o}_r}}{|\hat{\mathbf{u}}_r^{(p)}|}$ changes the module of $\hat{\mathbf{u}}_r^{(p)}$ by the observed value $\sqrt{\mathbf{o}_r}$ for all $r = 1, ...K$. In step 3 of the algorithm, the wave field at the first sensor plane is reconstructed by the wave field estimates from the Kth plane. $\mathbf{A}_{1,K}$ stands for the backward propagation operator from the Kth sensor plane to the first one. The algorithm updates cyclically the phase distributions at the observation planes. The reconstruction of the wave field is performed by the successive propagation from one sensor plane to another up to convergence.

Similar to [12], the object distribution has not appeared and has not been used in these iterations. In the successive algorithms, the reconstruction is concentrated on the sensor planes, where the observations are given. Contrary to it, the object wave field distribution is the main variable in the parallel algorithms updated in the iterations.

2. PROPOSED VARIATIONAL APPROACH

A typical variational setting for the Gaussian noise distribution in the observation model in Eq. (3) results in the following criterion:

$$J = \sum_{r=1}^{K} \frac{1}{2\sigma_r^2} ||\mathbf{o}_r - |\mathbf{u}_r|^2||_2^2 + \mu \cdot \mathrm{pen}(\mathbf{u}_0), \tag{11}$$

where the norm $||\cdot||_2^2$ is Euclidian.

The quadratic (fidelity) term in Eq. (11) appears due to the assumption that the noise in observations is Gaussian. The following term is the penalty (regularization) including prior information on the object distribution $\mathbf{u}_0$ to be reconstructed.

The criteria J_1 and J_2 discussed in Subsection 1.A are examples of different metrics that can be used as fidelity terms in variational formulations for wave field reconstruction. A more general criterion of the form $J_\gamma = \sum_r ||\mathbf{o}_r^{\gamma/2} - |\mathbf{u}_r|^\gamma||_2^2$, where γ is a parameter, is discussed in [20,21]. Mathematical aspects of the difference between this criterion with $\gamma = 1$ and $\gamma = 2$ are discussed in [5] and can be summarized as follows. For $\gamma = 1$, the criterion J_1 is not differentiable. It makes the mathematical analysis difficult and results in a lower convergence rate. For $\gamma > 1$, the criterion becomes differentiable and convenient for mathematical analysis. However, the experiments show that the accuracy for $\gamma = 1$ is better than that for $\gamma = 2$. As a compromise for $\gamma = 1$, the criterion J_1 can be modified to the form $J_{1,\delta} = \sum_r ||\mathbf{o}_r^{1/2} - |\mathbf{u}_r|^2 / \sqrt{|\mathbf{u}_r|^2 + \delta}||_2^2$, where $\delta > 0$ is small. $J_{1,\delta}$ is considered a differentiable approximation for J_1. More complex criteria are used for resolution-enhanced phase-retrieval methods with multiple

subarrays of sensors [21,22]. The corresponding criteria usually have a form of sums of the squared residuals for the subarrays.

In the criterion in Eq. (11), the choice in favor of $\gamma = 2$ is defined by the maximum likelihood approach and the assumption that the noise is i.i.d. Gaussian.

A. Constrained Maximum Likelihood

Based on the criterion in Eq. (11), we formulate the object wave field reconstruction as the following constrained optimization:

$$\hat{\mathbf{u}}_0 = \arg\min_{\mathbf{u}_0} \sum_{r=1}^{K} \frac{1}{2\sigma_r^2} ||\mathbf{o}_r - |\mathbf{u}_r|^2||_2^2 + \mu \cdot \mathrm{pen}(\mathbf{u}_0)$$

$$\text{subject to } \mathbf{u}_r = \mathbf{A}_r \mathbf{u}_0, \qquad r = 1, ..., K. \tag{12}$$

The regularization parameter μ in Eq. (12), as well as in Eq. (11), defines a balance between the accuracy of the observation fitting and a prior given by the penalty $\mathrm{pen}(\mathbf{u}_0)$. If $\mu = 0$, the solution $\hat{\mathbf{u}}_0$ minimizes $\sum_r \frac{1}{2\sigma_r^2} ||\mathbf{o}_r - |\mathbf{u}_r|^2||_2^2$, ignoring the fact that the data $\mathbf{o}_r$ are noisy. It can result in noisy and nonsmooth $\hat{\mathbf{u}}_0$. If $\mu > 0$ is comparatively large, then the noise effects are well suppressed but the solution $\hat{\mathbf{u}}_0$ can be oversmoothed, and important features lost. A proper selection of μ is an important point of the variation formulation.

Numerous forms of the penalty $\mathrm{pen}(\cdot)$ are used in literature on digital image processing derived from speculations varying from probabilistic modeling of image distribution priors to heuristic constructions (e.g., [23,24]).

In this paper, we use a simple quadratic Tikhonov's penalty [25] given in the form

$$\mathrm{pen}(\mathbf{u}_0) = ||\mathbf{u}_0||_2^2. \tag{13}$$

B. Augmented Lagrangian Method

By inserting Eq. (2) into Eq. (12), we can replace the constrained optimization by the unconstrained one:

$$\hat{\mathbf{u}}_0 = \arg\min_{\mathbf{u}_0} J,$$

$$J = \sum_{r=1}^{K} \frac{1}{2\sigma_r^2} ||\mathbf{o}_r - |\mathbf{A}_r \cdot \mathbf{u}_0|^2||_2^2 + \mu \cdot \mathrm{pen}(\mathbf{u}_0)\,. \tag{14}$$

Various gradient iterative algorithms (steepest descent, Newton, and Gauss–Newton) can be applied to find a solution of Eq. (14). The computational complexity of the gradient and second derivatives used in these algorithms is one of the main drawbacks of this unconstrained approach. Another principal drawback is the low convergence rate of these algorithms. It is recognized that a constrained optimization may lead to more efficient algorithms.

The augmented Lagrangian (AL) method, introduced independently by Hestenes [26] and Powell [27], is now standard for minimization in the presence of linear equality constraints. The AL criterion corresponding to Eq. (12) with complex-valued variables is of the form

$$\mathcal{L}(\mathbf{u}_0, \{\mathbf{u}_r\}, \{\Lambda_r\}) = \sum_{r=1}^{K} \frac{1}{\sigma_r^2} \left[\frac{1}{2} ||\mathbf{o}_r - |\mathbf{u}_r|^2||^2 + \frac{1}{\gamma_r} ||\mathbf{u}_r - \mathbf{A}_r \cdot \mathbf{u}_0||^2 + \frac{2}{\gamma_r} \mathrm{Re}\{\Lambda_r^H (\mathbf{u}_r - \mathbf{A}_r \cdot \mathbf{u}_0)\} \right] + \mu ||\mathbf{u}_0||_2^2, \tag{15}$$

where $\Lambda_r \in \mathbb{C}^n$ are the complex-valued Lagrange multipliers. The parameters γ_r are positive.

The Lagrangian based optimization is associated with the saddle problem, which requires minimization on $\mathbf{u}_0, \{\mathbf{u}_r\}$ and maximization on the vectors of the Lagrange multipliers $\{\Lambda_r\}$.

In Eq. (15), the linear constraints $\mathbf{u}_r - \mathbf{A}_r \cdot \mathbf{u}_0 = 0$ are used both in the linear and quadratic terms. If we keep only the quadratic terms, the augmented Lagrangian becomes the penalty criterion, which can be used assuming that the penalty coefficient $1/\gamma_r$ is large. As a rule, it leads to computational difficulties because this criterion can be very ill-conditioned. If we keep only the linear term, the AL becomes the standard Lagrangian. However, the saddle point of this standard Lagrangian is unstable. It may lead to problems with numerical calculations. The stability of the saddle point of AL is one of the principal advantages of this criterion.

The successive steps of AL optimization are as follows:

$$(\mathbf{u}_0^{t+1}, \{\mathbf{u}_r^{t+1}\}) \in \arg\min_{\mathbf{u}_0, \{\mathbf{u}_r\}} \mathcal{L}(\mathbf{u}_0, \{\mathbf{u}_r\}, \{\Lambda_r^t\}), \tag{16}$$

$$\Lambda_r^{t+1} = \Lambda_r^t + \alpha_r \cdot (\mathbf{u}_r^{t+1} - \mathbf{A}_r \cdot \mathbf{u}_0^t), \qquad r = 1, ..., K. \tag{17}$$

Note that, in Eq. (16), the minimization on $\mathbf{u}_0, \{\mathbf{u}_r\}$ is produced for fixed $\{\Lambda_r^t\}$. The steps on the Lagrangian multipliers [Eq. (17)] are produced in the gradient direction (according to the minimum condition $\nabla_{\Lambda_r^*} L = 0$) with the step size α_r.

Alternating direction multiplier methods have been extensively developed to minimize the AL criteria. In these methods, the optimization variables (in our case $\mathbf{u}_0$, $\{\mathbf{u}_r\}$) are partitioned into several blocks according to their roles, and then the augmented Lagrangian function is minimized with respect to each block by fixing all other blocks at each inner iteration [28]. It leads to the alternating minimization on $\mathbf{u}_0$, $\{\mathbf{u}_r\}$, and results in the algorithm

$$\text{For } t = 0, 1, ...,$$

$$\text{For } r = 1, ...K,$$

$$\mathbf{u}_r^{t+1} \in \arg\min_{\{\mathbf{u}_r\}} \mathcal{L}(\mathbf{u}_0^t, \{\mathbf{u}_r\}, \{\Lambda_r^t\}), \tag{18}$$

$$\Lambda_r^{t+1} = \Lambda_r^t + \alpha_r \cdot (\mathbf{u}_r^{t+1} - \mathbf{A}_r \cdot \mathbf{u}_0^t),$$

$$\text{End on } r, \tag{19}$$

$$\mathbf{u}_0^{t+1} \in \arg\min_{\mathbf{u}_0} \mathcal{L}(\mathbf{u}_0, \{\mathbf{u}_r^{t+1}\}, \{\Lambda_r^t\}), \qquad \text{End on } t. \tag{20}$$

This type of algorithm has recently become the subject of intensive development and study, in particular, for image processing. These algorithms, sharing many common ideas and features, appear under different names, such as split Bregman iterations [29], iterative shrinkage-thresholding algorithms [30], alternating direction method of multipliers [31], and majorization–minimization algorithms [32].

C. Proposed Algorithm

In order to derive the reconstruction algorithm, we minimize $\mathcal{L}$ on $\{\mathbf{u}_r\}$ and $\mathbf{u}_0$. Details of the corresponding calculations are given in Appendix A. These minimizations lead to the following iterative AL algorithms:

AL Algorithm

1. Set $t=0$ (initialization), $\mathbf{u}_0^0$, Λ_r^0,
2. Repeat for $t=0,1,...,$
3. Repeat for $r=1,...K$,
4. $\mathbf{u}_r^{t+1/2}=\mathbf{A}_r\cdot\mathbf{u}_0^t$,
5. $\mathbf{u}_r^{t+1}[k]=\mathcal{G}(\mathbf{o}_r[k],\quad \mathbf{u}_r^{t+1/2}[k],\Lambda_r^t[k])$,
6. $\Lambda_r^{t+1}=\Lambda_r^t+\alpha_r\cdot(\mathbf{u}_r^{t+1}-\mathbf{u}_r^{t+1/2})$,
7. End on r,
8. $$\mathbf{u}_0^{t+1}=\left(\sum_{r=1}^{K}\frac{1}{\gamma_r\sigma_r^2}\mathbf{A}_r^H\mathbf{A}_r+\mu\cdot\mathbf{I}_{n\times n}\right)^{-1}\times\sum_{r=1}^{K}\frac{1}{\gamma_r\sigma_r^2}\mathbf{A}_r^H(\mathbf{u}_r^{t+1}+\Lambda_r^t),$$
9. End on t. (21)

The initialization for $t=0$ concerns the object plane distribution $\mathbf{u}_0^0$ and Lagrange multipliers Λ_r^0. Step 4 returns $\mathbf{u}_r^{t+1/2}$ according to the forward propagation of the object wave field $\mathbf{u}_0^t$ to the sensor planes. Step 5 returns the updates $\mathbf{u}_r^{t+1}$ of the wave field distributions at the sensor planes by fitting $\mathbf{u}_r^{t+1/2}$ with the observation $\mathbf{o}_r$. The operator defining this update is denoted in the algorithm as $\mathcal{G}$. The formulas defining the corresponding calculations in Eqs. (A5) and (A6) are given in Appendix A. Step 6 returns the updated Lagrange multipliers. Step 8 gives the update for the object wave field calculated from the found estimates $\mathbf{u}_r^{t+1}$ and the Lagrange multipliers Λ_r^t.

Note that Step 8 has a structure that is typical for the parallel algorithms from Subsection 1.A. The object wave field reconstruction is calculated using the summation of the object estimates obtained using the backward propagation of $\mathbf{u}_r^{t+1}$. The backward propagation operator is of the form $\mathbf{B}^{-1}\mathbf{A}_r^H$, $\mathbf{B}=\sum_{r=1}^{K}\mathbf{A}_r^H\mathbf{A}_r+\mu\cdot\mathbf{I}_{n\times n}$. In Step 8, the multiple estimates of $\mathbf{u}_0$ obtained from the multiple sensor planes are aggregated into the final update for the object reconstruction. This aggregation formula is similar to Eq. (9).

The proposed algorithm is derived and introduced using the vector-matrix notation. The computational complexity of the algorithm in this form is reasonable only for images that are small, because the matrices $\mathbf{A}_r$ are of the size $n\times n$, $n=N\cdot M$. In order to make the algorithm applicable for large size images, we implement the algorithm using fast Fourier transform (FFT) calculations.

First, note that Eq. (2) is a matrix representation for the convolution of the object distribution u_0 (with the finite rectangular support $N\times M$) and the shift-invariant kernel of the wave field propagation operator (with infinite support). For the exact calculations of the convolution in the FFT domain, we need to use the double size version of the propagation kernels, with the support $2N\times 2M$, and the object wave field $\tilde{u}_0$ zero-padded to the same size $2N\times 2M$ (see [9]).

Further, in the FFT double size manipulations, the variable $\tilde{u}_r$ is calculated for the double size support $2N\times 2M$, where the central part of this support corresponds to the sensor of the original size $N\times M$. In this way, at the sensor planes, we have additional wave field estimates for the areas outer to the sensor supports. We use these extra estimates at the sensor planes to improve the performance of the algorithm.

Details concerning this sort of fruitful use of the extra estimates appearing due to double size FFT calculations can be seen in [11], where they are exploited in the inverse wave field reconstruction based on F-DDT propagation modeling.

3. SIMULATION EXPERIMENTS

The main goal of the numerical experiments is to analyze the accuracy of wave field reconstruction and to study the performance of the algorithm for different types of object distributions, different parameters $\{K,z_f,\Delta_z\}$ of the optical setup, and different parameters $\{\gamma_r,\alpha_r,\mu\}$ of the algorithm.

We consider the complex-valued object distribution in the form $u_0[k]=|u_0[k]|\cdot\exp(j\cdot\phi_0[k])$. For the object with amplitude modulation (AM), $\phi_0[k]\equiv 0$, $|u_0[k]|=w[k]$. For the object with phase modulation (PM), $|u_0[k]|\equiv 1$, $\phi_0[k]=\pi(w[k]-\frac{1}{2})$. Here $w[k]$ is a spatially varying test image. It is assumed that $0.1\le w[k]\le 1$.

The following set of square ($N\times N$) test images is used: gray-scale *lena* (256×256), binary *logo* (256×256), *chessboard* (128×128), and the gray-scale smooth *Mexican Hat* (200×200). The last test image is defined by the formula $w[k]=-\frac{4}{3\sqrt{\pi}}\cdot(||\frac{k}{40}||^2-2)\cdot\exp(-\frac{1}{2}||\frac{k}{40}||^2)$ and is calculated on the 2D integer grid $-100\le(k_1,k_2)\le 99$. Thus $w[k]$ is normalized in such a way that $0.1\le w[k]\le 1$.

The accuracy of the wave field reconstruction is characterized by the root-mean-square error (RMSE) criterion calculated for amplitude and phase of the wave field.

The phase from the intensity measurements can be reconstructed up to an arbitrary constant only. In order to eliminate this ambiguity, RMSE for the object phase is calculated for $\phi_0-\hat{\phi}_0-\mathrm{mean}(\phi_0-\hat{\phi}_0)$, where $\hat{\phi}_0$ is an estimate of the phase, and $\mathrm{mean}(\phi_0-\hat{\phi}_0)$ stands for the mean value of the estimation error calculated over a test image.

The pixelated models for the object and sensor planes have square pixels $\Delta\times\Delta$, $\Delta=6.7\,\mu$m with 100% fill factors. The wavelength $\Lambda=532$ nm corresponding to an Nd:YAG green laser. The "in-focus" distance for the considered lensless scenario is calculated as $z_f=N\cdot\Delta^2/\Lambda$ (see [33]). The distance from the object plane to the first sensor plane z_1 is expressed through this "in-focus" distance as $z_1=d\cdot z_f$, where the parameter d is varying in the interval $[0.5,3]$. The number of measurement planes K takes values from the interval [3,10].

If it is not specified, we use $d=2$, $K=5$, and $\Delta_z=2$ mm.

The observations $\mathbf{o}_r$ are generated from $\mathbf{u}_0$ using the double size F-DDT. The observation fields generated in this way are precise for pixelated sensor and object distributions [11]. This F-DDT is exploited also in the AL algorithm for the forward propagation operators $\mathbf{A}_r$ [Step 4 and Step 8 in Eq. (21)].

The MATLAB code used for the simulation experiments, the results, and the discussions are available on the website [34].

A. Parameters of the AL Algorithm

The performance of the algorithm depends essentially on the parameters of the algorithm. In this subsection, we present the values of these parameters, which are acceptable for various scenarios and for the considered test images. For simplicity, we assume that $\alpha_r=\alpha$, $\gamma_r=\gamma$, $\sigma_r^2=\sigma^2$ for all r. Then, Step 8 of the algorithm in Eq. (21) can be rewritten as

$$\mathbf{u}_0^{t+1}=\left[\sum_{r=1}^{K}\mathbf{A}_r^H\mathbf{A}_r+\tilde{\mu}\gamma\cdot\mathbf{I}_{n\times n}\right]^{-1}\sum_{r=1}^{K}\mathbf{A}_r^H(\mathbf{u}_r^{t+1}+\Lambda_r^t),\qquad(22)$$

where the regularization parameter μ is changed for $\tilde{\mu} = \mu \cdot \sigma^2$. Thus, we are looking for the best choice of the triplet of the AL parameters $\{\gamma, \alpha, \tilde{\mu}\}$ that works for the considered test images and the various parameters of the optical setup.

The value of the pixel size Δ is crucial for the algorithm's performance. The optimal values of $\tilde{\mu}$ and γ may differ in several times depending on Δ, especially for PM. For a smaller Δ, the optimal $\tilde{\mu}$ is larger and the optimal γ is smaller.

The results on optimization of the algorithm parameters can be summarized as follows:

a. the recommended penalty coefficient $\gamma = 10$;
b. the step-size parameter for the Lagrange multipliers $\alpha = 1$; and
c. the regularization parameter $\tilde{\mu}$ is about 10 times larger for PM than for AM. For noiseless data ($\sigma = 0$), we use $\tilde{\mu} = 5 \cdot 10^{-4}/10^{-3}$ for AM/PM, respectively. For noisy data, provided that the level of the noise is low ($\sigma \in [0.01, 0.05]$), we use $\tilde{\mu} = 5 \cdot 10^{-3}/10^{-2}$ for AM/PM, respectively. Larger values of the regularization parameter $\tilde{\mu}$ are recommended for larger d and for larger K.

For each experiment, the reconstruction accuracy can be improved by modifying $\{\gamma, \alpha, \tilde{\mu}\}$. More comments concerning the parameters of the algorithm are presented in [34].

B. Quality and Accuracy of Imaging

Qualitative and quantitative performance of the algorithm is considered in this subsection for noiseless ($\sigma = 0$) and noisy data ($\sigma = 0.05$) with the parameters $\{\gamma, \alpha, \tilde{\mu}\}$ fixed as it is discussed above.

In Fig. 2, we demonstrate the algorithm performance depending on the number of measurement planes K and on distances z_1. The amplitude reconstructions for the AM object with the binary test image logo are shown. The columns in Fig. 2 correspond to different numbers of the observation planes (from the left to the right) $K = 3, 5, 10$.. The rows are given for different z_1 varying from $0.5 \cdot z_f$ to $3 \cdot z_f$ (from top to bottom), $z_f = 21.6\,\text{mm}$. The distances between the sensor planes are fixed, $\Delta_z = 2\,\text{mm}$.

It can be seen that the larger K leads to a monotonically better quality of the wave field reconstruction for different z_1 smaller or larger than z_f.

The relative improvement is especially valuable for small K (about 30% in RMSE values for two additional planes from $K = 3$ to $K = 5$) and not so essential for larger K (from $K = 5$ to $K = 10$).

It is seen that, for $z_1 = 0.5 \cdot z_f$, a very good result is obtained already for $K = 5$ (the second image in the top row, $\text{RMSE}(|u_0|) = 0.0106$). Here $z_5 \approx 18.8\,\text{mm} < z_f$. The accuracy and imaging for $K = 10$ are slightly better (the third image in the top row, $\text{RMSE}(|u_0|) = 0.0101$), even though some measurement planes are far from the focus distance, e.g., $z_7 \approx 22.8\,\text{mm} > z_f$. For other rows, larger K give more noticeable improvements in imaging (in these cases, $z_r \geq z_f$ for all r). The corresponding RMSE values are presented in Table 1.

In Fig. 3, we consider the accuracy of the phase reconstruction depending on the distance between the sensor planes Δ_z. These results are presented for different K, image sizes N, and distances z_1. The curves are RMSE values calculated for the PM object with the binary test image chessboard, $\sigma = 0.05$.

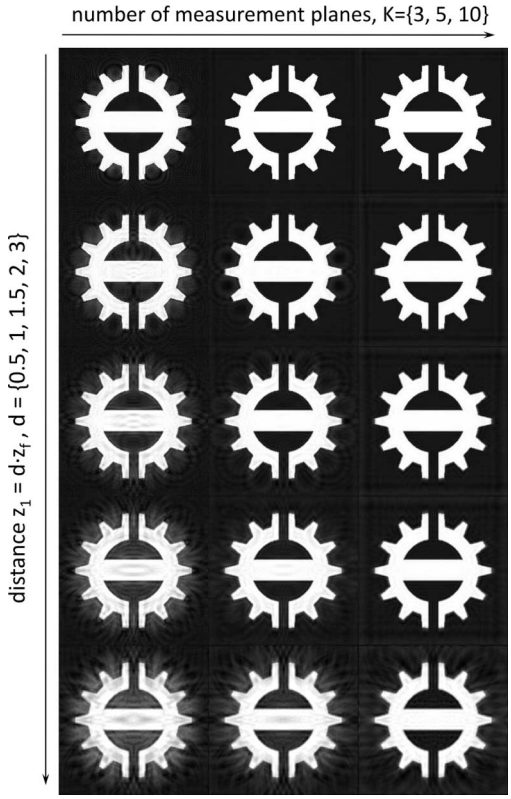

Fig. 2. Reconstructions of the amplitude for the AM object distribution with different numbers of sensor planes K and distances z_1 between the object and the first sensor plane. Logo test-image, noiseless data $\sigma = 0$, $\tilde{\mu} = 5 \cdot 10^{-3}$.

All the RMSE curves go rapidly down, achieve minimum values corresponding to optimization on Δ_z, and then grow initially slowly and then faster for $\Delta_z > 0.5 \cdot z_f$. The existence of the optimal Δ_z is well seen. It can be noticed from comparison of the curves that a larger number of the observation planes K, a larger image size N, and a smaller distance z_1 lead to a proportional decrease of the optimal value of Δ_z. However, the RMSE curves are quite irregular and multimodal, and it is not clear how to select the best Δ_z. Note that, despite a clear decrease of RMSE for larger Δ_z, the influence of this parameter on the quality of imaging is not essential.

There are interesting connections between the accuracy of the AL algorithm and the condition number of the operator $\sum_{r=1}^{K} \mathbf{A}_r^H \mathbf{A}_r$ in Eq. (22). Smaller (larger) values of this condition number result in a higher (lower) accuracy of the

Table 1. Quantitative Comparison of the Amplitude Reconstruction, $\text{RMSE}(|u_0|)$ for the Test, Presented in Fig. 2

	K		
d	3	5	10
0.5	0.041	0.0106	**0.0101**
1	0.0691	0.0371	0.0323
1.5	0.0882	0.0597	0.0512
2	0.1159	0.072	0.0628
3	0.1478	0.1148	0.091

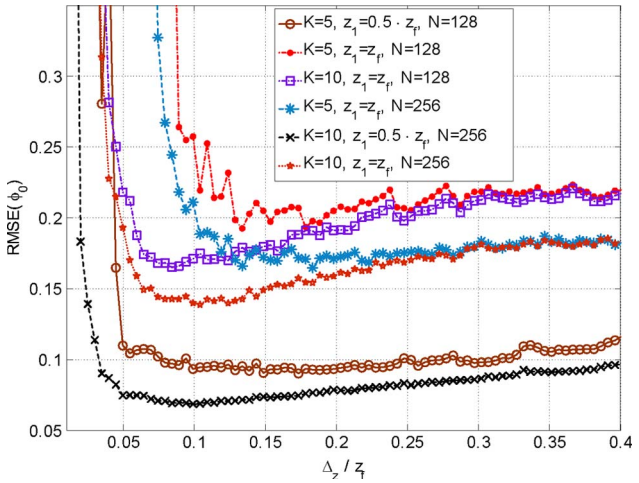

Fig. 3. (Color online) Accuracy of the phase reconstruction with respect to the distance between measurement planes Δ_z. PM object, chessboard test image, noisy data $\sigma = 0.05$.

reconstruction. The condition number can be used as a criterion for selection of the parameters $\{K, z_f, \Delta_z\}$ of the optical setup. Details of this topic are beyond the scope of this paper. Some analysis and comments can be seen in [34].

C. Comparison of AL with SBMIR

In this subsection, we compare of the performance and the accuracies of the proposed AL algorithm in Eq. (21) with respect to the SBMIR algorithm in Eq. (10).

First, in Fig. 4, we present the reconstruction of the object amplitude and phase for the AM object with the gray-scale lena as the test image w. These reconstructions are done for the noiseless case of $\sigma = 0$. The first column is obtained by AL: Fig. 4(a) is the reconstruction of amplitude, $\mathrm{RMSE}(|u_0|) = 0.041$, and Fig. 4(c) is the reconstruction of phase, $\mathrm{RMSE}(\phi_0) = 0.091$. The second column images are obtained by SBMIR: Fig. 4(b) is for the amplitude reconstruction,

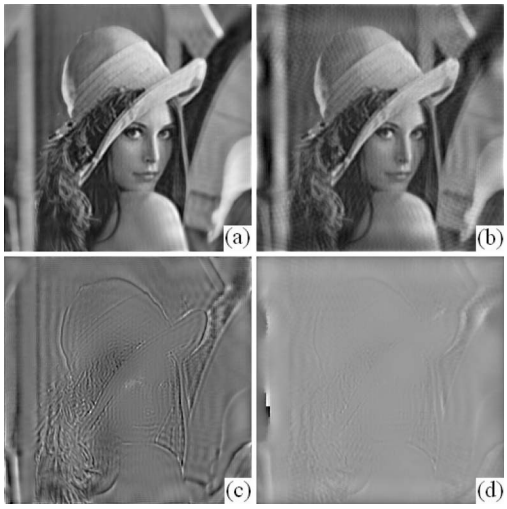

Fig. 4. Comparison of the amplitude and phase reconstructions obtained by AL (left column) and SBMIR (right column) algorithms. The top row demonstrates the amplitude reconstructions (a) by AL, $\mathrm{RMSE}(|u_0|) = 0.041$, and (b) by SBMIR, $\mathrm{RMSE}(|u_0|) = 0.08$. The bottom row shows the phase reconstructions, obtained (c) by AL, $\mathrm{RMSE}(\phi_0) = 0.091$, and (d) by SBMIR, $\mathrm{RMSE}(\phi_0) = 0.28$. AM object, lena test image, $K = 5$, noiseless data $\sigma = 0$.

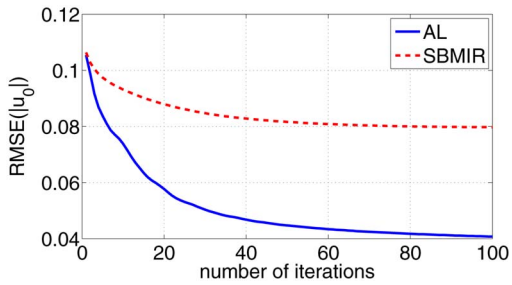

Fig. 5. (Color online) Convergence rates of the AL and SBMIR algorithms for the test presented in Fig. 4.

$\mathrm{RMSE}(|u_0|) = 0.08$, and Fig. 4(d) is for the phase reconstruction, $\mathrm{RMSE}(\phi_0) = 0.28$.

The accuracy of the AL algorithm for both the amplitude and phase reconstructions is approximately twice better than that for the SBMIR algorithm. The advantage of the AL algorithm for the amplitude reconstruction is obvious because the SBMIR reconstruction of the amplitude is corrupted by artifacts parallel to the image borders. There are no such artifacts in the AL imaging.

The convergence rates of the AL and SBMIR algorithms for this experiment are shown in Fig. 5. The proposed AL algorithm demonstrates not only a significantly better reconstruction accuracy but also a good convergence rate.

In Fig. 6 we compare the wave field reconstructions for the PM object distribution with the binary test image chessboard used for the phase modulation. The results are shown for the noisy data, $\sigma = 0.05$. The left column of Fig. 6 is obtained by AL: Fig. 6(a) shows the amplitude reconstruction, $\mathrm{RMSE}(|u_0|) = 0.23$, and Fig. 6(c) shows the phase reconstruction, $\mathrm{RMSE}(\phi_0) = 0.26$. The right column in Fig. 6 is obtained by SBMIR: Fig. 6(b) is for the amplitude reconstruction, $\mathrm{RMSE}(|u_0|) = 0.35$, and Fig. 6(d) is for the phase reconstruction, $\mathrm{RMSE}(\phi_0) = 0.58$.

The AL algorithm demonstrates clear and sharp phase imaging, while the SBMIR phase reconstruction is blurred

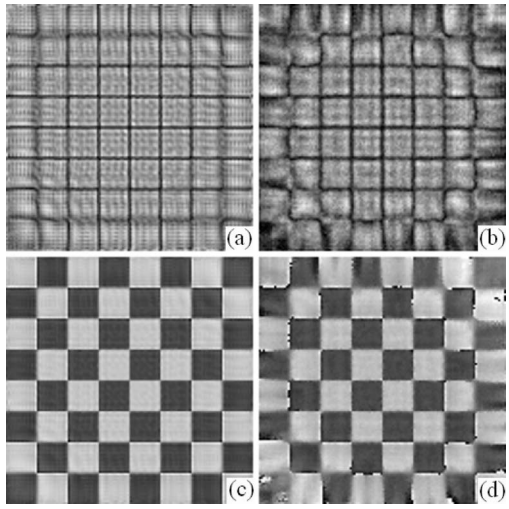

Fig. 6. Comparison of the amplitude and phase reconstructions obtained by AL (left column) and SBMIR (right column) algorithms. The top row demonstrates the amplitude reconstructions (a) by AL, $\mathrm{RMSE}(|u_0|) = 0.23$, and (b) by SBMIR, $\mathrm{RMSE}(|u_0|) = 0.35$. The bottom row illustrates the phase reconstructions, obtained (c) by AL, $\mathrm{RMSE}(\phi_0) = 0.26$, and (d) by SBMIR, $\mathrm{RMSE}(\phi_0) = 0.58$. PM object, chessboard test image, $K = 5$, noisy data $\sigma = 0.05$.

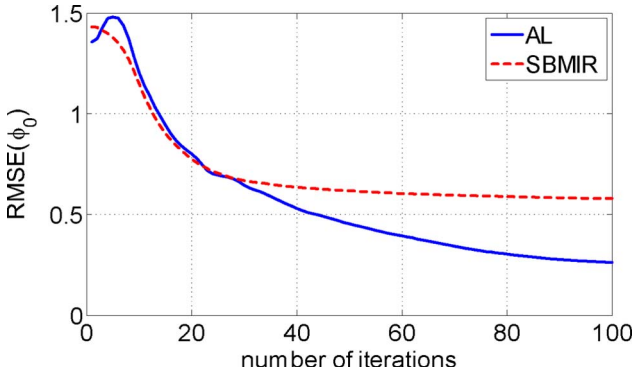

Fig. 7. (Color online) Convergence rates of the AL and SBMIR algorithms for the test presented in Fig. 6.

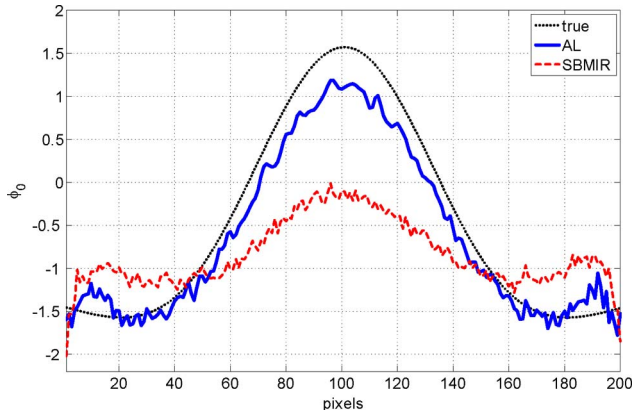

Fig. 8. (Color online) Cross sections of the true phase (dotted curve) and phase reconstructions obtained for the Mexican Hat test image. The solid curve corresponds to AL, RMSE(ϕ_0) = 0.187, and the dashed curve corresponds to SBMIR, RMSE(ϕ_0) = 0.511. PM object, 1000 iterations, $K = 5$, noisy data $\sigma = 0.05$.

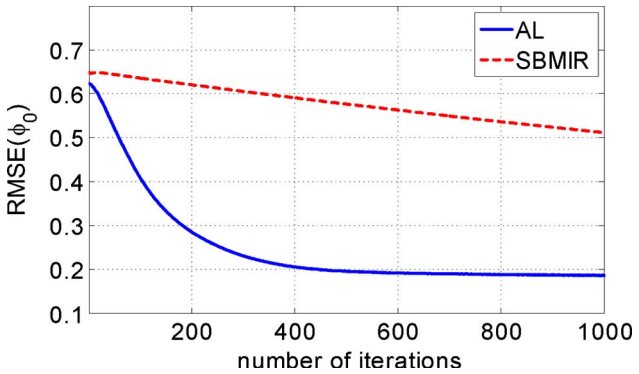

Fig. 9. (Color online) Convergence rates of the AL and SBMIR algorithms for the phase reconstruction (Mexican Hat) of the test presented in Fig. 8.

and significantly destroyed. The RMSE values for the amplitude $|u_0|$ and phase ϕ_0 reconstructions are large because of the large errors on the borders of the blocks of the chessboard. Overall, the reconstruction accuracy both for amplitude and for phase is much better for the AL algorithm.

The convergence rates of the AL and SBMIR algorithms for the phase reconstruction are compared in Fig. 7. After the first 20 iterations, where the accuracies of the algorithms are equivalent, the convergence of SBMIR becomes slow and, after 100 iterations, AL shows a significant advantage in the achieved RMSE values.

In contrast to the discontinuous and nonsmooth distributions (lena, chessboard, and logo test images considered above), the test image Mexican Hat is smooth. It is used for the phase modulation of the object distribution. In Fig. 8 we show the cross sections of the true phase and its normalized reconstructions, obtained by the AL and SBMIR algorithms. The advantage of the AL algorithm is obvious. The almost complete bell and the hollow of the Mexican Hat are reconstructed by AL (solid curve in Fig. 8). Larger errors appear on the borders in the AL and SBMIR reconstructions.

The convergence rates of the AL and SBMIR algorithms for this PM experiment are illustrated in Fig. 9. The AL algorithm converges much faster than SBMIR and gives essentially better RMSE values.

D. Computational Complexity

In Table 2, we show the average computational time (in seconds) for 100 iteration of the AL algorithm depending on the number of the measurement planes K and the image size defined by the parameter N. In our experiments, the image size in pixels is $n = N^2$.

These numbers are obtained by Monte Carlo simulations with averaging over 50 experiments. The computer used for experiments is Intel Core 2Duo E8400 at 3 GHz; RAM, 4 GB; Windows Xp SP3; MATLAB 7.9.0 (R2009b). Memory requirements depend mainly on N and K.

The averaged computational times required for calculation of $\{\mathbf{A}_r\}_{r=1}^{K}$ in the algorithm and more comments about the computational complexity are presented on our web page [34].

4. CONCLUSION

In this work, we present a novel variational formulation for the phase-retrieval problem. Being the maximum likelihood style, this setting takes into consideration the Gaussian noise distribution. Based on the AL technique, we have developed the iterative constrained optimization algorithm for the amplitude and phase reconstruction assuming that the object distribution is complex valued.

Table 2. Computational Time (in Seconds) for 100 Iterations of the AL Algorithm

	K					
N	2	3	5	7	10	15
128	10.4	15.6	24.8	27	38.2	56.3
256	46.2	55.6	88.1	121.7	172.9	257.6
512	180.6	253.9	398.5	556.9	787.5	1176.2

The algorithm demonstrates a good convergence rate, and good accuracy and imaging for phase and amplitude of the wave field reconstruction both for noiseless and noisy observation data.

APPENDIX A: DERIVATION OF THE AL ALGORITHM

Here we provide some details of minimization of $\mathcal{L}$ on $\mathbf{u}_0$ on $\{\mathbf{u}_r\}$ defining the AL algorithm in Eq. (21).

1. Optimization on $\mathbf{u}_r$

Let us rewrite the criterion of Eq. (15) into the form

$$\mathcal{L}(\mathbf{u}_0,\{\mathbf{u}_r\},\{\Lambda_r\})=\sum_{r=1}^{K}\frac{1}{\sigma_r^2}\left\{\frac{1}{2}||\mathbf{o}_r-|\mathbf{u}_r|^2||^2+\frac{1}{\gamma_r}||\mathbf{u}_r-\mathbf{A}_r\cdot\mathbf{u}_0||^2+\frac{1}{\gamma_r}\left[\Lambda_r^H(\mathbf{u}_r-\mathbf{A}_r\cdot\mathbf{u}_0)+(\mathbf{u}_r-\mathbf{A}_r\cdot\mathbf{u}_0)^H\Lambda_r\right]\right\}+\mu||\mathbf{u}_0||_2^2. \tag{A1}$$

The Lagrangian Eq. (A1) is additive with respect to the vectors $\mathbf{u}_r$ and their components. Thus, the minimization on $\mathbf{u}_r$ can be produced in the elementwise manner. The derivative $\nabla_{\mathbf{u}_r^*[k]}\mathcal{L}=0$ gives the minimum condition for $\mathbf{u}_r[k]$ as

$$\frac{\partial L}{\partial\mathbf{u}_r^*[k]}=\frac{1}{\sigma_r^2}(|\mathbf{u}_r[k]|^2-\mathbf{o}_r[k])\cdot\mathbf{u}_r[k]+\frac{1}{\gamma_r\sigma_r^2}(\mathbf{u}_r[k]-(\mathbf{A}_r\cdot\mathbf{u}_0)[k]+\Lambda_r[k])=0. \tag{A2}$$

It follows that

$$\mathbf{u}_r[k]=\frac{(\mathbf{A}_r\cdot\mathbf{u}_0)[k]-\Lambda_r[k]}{\gamma_r(|\mathbf{u}_r[k]|^2-\mathbf{o}_r[k])+1}=\frac{\eta_r[k]}{\kappa_r[k]}. \tag{A3}$$

Taking the module from the left and right sides of Eq. (A3), we arrive at the cubic equation with respect to $|\mathbf{u}_r[k]|$:

$$|\mathbf{u}_r[k]|^3+|\mathbf{u}_r[k]|\cdot\left(\frac{1}{\gamma_r}-\mathbf{o}_r[k]\right)-\mathrm{sgn}(\kappa_r[k])\cdot\frac{|\eta_r[k]|}{\gamma_r}=0. \tag{A4}$$

In reality, we have two different cubic equations: corresponding to $\mathrm{sgn}(\kappa_r[k])=1$ and to $\mathrm{sgn}(\kappa_r[k])=-1$. Each of these Eqs. (A4) may have a single or three real solutions. We are looking for a nonnegative real root denoted as $|\tilde{\mathbf{u}}_r[k]|$. It can be seen that such $|\tilde{\mathbf{u}}_r[k]|$ always exists.

If this root is found, the corresponding complex-valued estimate of the wave field at the sensor plane $\hat{\mathbf{u}}_r[k]$ is calculated according to Eq. (A3) as

$$\hat{\mathbf{u}}_r[k]=\frac{(\mathbf{A}_r\cdot\mathbf{u}_0)[k]-\Lambda_r[k]}{\gamma_k(|\tilde{\mathbf{u}}_r[k]|^2-\mathbf{o}_r[k])+1}. \tag{A5}$$

The nonnegative solution $|\tilde{\mathbf{u}}_r[k]|$ can be not unique. Then the procedure becomes more complex because these solutions should be analyzed. As a proper solution, we selected $\hat{\mathbf{u}}_r[k]$ [Eq. (A5)] giving the minimum value to the corresponding rth summand of $\mathcal{L}$.

We denote the nonlinear algorithm giving $\hat{\mathbf{u}}_r[k]$ defined by Eq. (A5) and including the mentioned analysis of the multiple nonnegative solutions as

$$\hat{\mathbf{u}}_r[k]=\mathcal{G}(\mathbf{o}_r[k],\mathbf{u}_r[k],\Lambda_r[k]), \tag{A6}$$

where $\mathbf{u}_r[k]=(\mathbf{A}_r\cdot\mathbf{u}_0)[k]$.

It defines Step 5 of the algorithm in Eq. (21).

2. Optimization on $\mathbf{u}_0$

Minimization on a complex-valued $\mathbf{u}_0=\mathrm{Re}\{\mathbf{u}_0\}+i\mathrm{Im}\{\mathbf{u}_0\}$ ($\mathbf{u}_0=\mathbf{a}_0+i\mathbf{b}_0$) means minimization on both the real $\mathbf{a}_0$ and imaginary $\mathbf{b}_0$ parts of $\mathbf{u}_0$. The necessary minimum conditions would have the standard form $\partial\mathcal{L}/\partial\mathbf{a}_0=0$ and $\partial\mathcal{L}/\partial\mathbf{b}_0=0$. However, it is more convenient (and the resulting equations are more compact) if one replaces these real and imaginary parts by the complex-valued $\mathbf{u}_0$ and $\mathbf{u}_0^*$, where the $*$ superscript denotes the complex conjugate. Then, the necessary minimum conditions have the form $\partial\mathcal{L}/\partial\mathbf{u}_0^*=0$ or, equivalently, $\partial\mathcal{L}/\partial\mathbf{u}_0=0$. In these derivative calculations, the variables $\mathbf{u}_0$ and $\mathbf{u}_0^*$ are treated as independent. Note also that the differentiation of a scalar by a vector results in a vector of derivatives: $\partial\mathcal{L}/\partial\mathbf{u}_0=[\partial L/\partial\mathbf{u}_0[1],\ldots,\partial L/\partial\mathbf{u}_0[n]]^T$, provided that $\mathbf{u}_0=[\mathbf{u}_0[1],\ldots,\mathbf{u}_0[n]]^T$. In particular, for $||\mathbf{u}_0||_2^2=\mathbf{u}_0^T\mathbf{u}_0^*$, we have $\partial||\mathbf{u}_0||_2^2/\partial\mathbf{u}_0^*=\mathbf{u}_0$ and $\partial||\mathbf{u}_0||_2^2/\partial\mathbf{u}_0=\mathbf{u}_0^*$.

The minimum condition for Eq. (A1) in the form $\nabla_{\mathbf{u}_0^*}L=0$ gives the solution in the form

$$\hat{\mathbf{u}}_0=\left(\sum_{r=1}^{K}\frac{1}{\gamma_r\sigma_r^2}\mathbf{A}_r^H\mathbf{A}_r+\mu\cdot\mathbf{I}_{n\times n}\right)^{-1}\sum_{r=1}^{K}\frac{1}{\gamma_r\sigma_r^2}\mathbf{A}_r^H(\mathbf{u}_r+\Lambda_r). \tag{A7}$$

It defines Step 8 of the AL algorithm.

3. Optimization on Λ_r

The update of the Lagrangian multipliers in the gradient direction ($\nabla_{\Lambda_r^*}\mathcal{L}=\mathbf{u}_r-\mathbf{A}_r\cdot\mathbf{u}_0$) is defined by the equation

$$\Lambda_r^{t+1}=\Lambda_r^t+\alpha_r\cdot(\mathbf{u}_r^{t+1}-\mathbf{A}_r\cdot\mathbf{u}_0^t). \tag{A8}$$

It defines Step 6 of the algorithm.

ACKNOWLEDGMENTS

This work was supported by the Academy of Finland, project no. 213462, 2006-2011 (Finnish Programme for Centres of Excellence in Research) and project no. 138207, 2011-2014. The postgraduate work of Artem Migukin is funded by Tampere Doctoral Programme in Information Science and Engineering (TISE).

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. B. Gu and G. Yang, "On the phase retrieval problem in optical and electronic microscopy," Acta Opt. Sin. **1**, 517–522 (1981).
3. 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).
4. J. R. Fienup, "Phase retrieval algorithms: a comparison," Appl. Opt. **21**, 2758–2769 (1982).
5. D. R. Luke, J. V. Burke, and R. Lyon, "Optical wavefront reconstruction: theory and numerical methods," SIAM Rev. **44**, 169–224 (2002).

6. J. Burke and D. R. Luke, "Variational analysis applied to the problem of optical phase retrieval," SIAM J. Control Optim. **42**, 576–595 (2003).
7. H. H. Bauschke, P. L. Combettes, and D. R. Luke, "Phase retrieval, error reduction algorithm, and Fienup variants: a view from convex optimization," J. Opt. Soc. Am. A **19**, 1334–1345 (2002).
8. J. W. Goodman, *Introduction to Fourier Optics*, 2nd ed. (McGraw-Hill, 1996).
9. F. Shen and A. Wang, "Fast-Fourier-transform based numerical integration method for the Rayleigh-Sommerfeld diffraction formula," Appl. Opt. **45**, 1102–1110 (2006).
10. V. Katkovnik, A. Migukin, and J. Astola, "Backward discrete wave field propagation modeling as an inverse problem: toward perfect reconstruction of wave field distributions," Appl. Opt. **48**, 3407–3423 (2009).
11. V. Katkovnik, J. Astola, and K. Egiazarian, "Discrete diffraction transform for propagation, reconstruction, and design of wavefield distributions," Appl. Opt. **47**, 3481–3493 (2008).
12. D. L. Misell, "A method for the solution of the phase problem in electron microscopy," J. Phys. D **6**, L6–L9 (1973).
13. T. Isernia, G. Leone, R. Pierri, and F. Soldovieri, "The role of support information and zero locations in phase retrieval by a quadratic approach," J. Opt. Soc. Am. A **16**, 1845–1856 (1999).
14. R. W. Deming, "Phase retrieval from intensity-only data by relative entropy minimization," J. Opt. Soc. Am. A **24**, 3666–3679 (2007).
15. F. Soldovieri, R. Deming, and R. Pierri, "An improved version of the relative entropy minimization approach for the phase retrieval problem," AEU Int. J. Electron. Commun. **64**, 56–65 (2010).
16. A. Migukin, V. Katkovnik, and J. Astola, "Multiple plane phase retrieval based on inverse regularized imaging and discrete diffraction transform," AIP Conf. Proc. **1236**, 81–86 (2010).
17. 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).
18. P. Almoro, G. Pedrini, and W. Osten, "Complete wavefront reconstruction using sequential intensity measurements of a volume speckle field," Appl. Opt. **45**, 8596–8605 (2006).
19. P. Almoro, A. M. Maallo, and S. Hanson, "Fast-convergent algorithm for speckle-based phase retrieval and a design for dynamic wavefront sensing," Appl. Opt. **48**, 1485–1493 (2009).
20. M. Guizar-Sicairos and J. R. Fienup, "Phase retrieval with transverse translation diversity: a nonlinear optimization approach," Opt. Express **16**, 7264–7278 (2008).
21. X. Hu, S. Li, and Y. Wu, "Resolution-enhanced subpixel phase retrieval method," Appl. Opt. **47**, 6079–6087 (2008).
22. G. R. Brady, M. Guizar-Sicairos, and J. R. Fienup, "Optical wavefront measurement using phase retrieval with transverse translation diversity," Opt. Express **17**, 624–639 (2009).
23. L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," Physica D (Amsterdam) **60**, 259–268 (1992).
24. M. Bertero and P. Boccacci, *Introduction to Inverse Problems in Imaging* (IOP, 1998).
25. A. N. Tikhonov and V. Y. Arsenin, *Solution of Ill-Posed Problems* (Wiley, 1977).
26. M. R. Hestenes, "Multiplier and gradient methods," J. Optim. Theory Appl. **4**, 303–320 (1969).
27. M. J. D. Powell, "A method for nonlinear constraints in minimization problems," in *Optimization*, R. Fletcher, ed. (Academic, 1969), pp. 283–298.
28. D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. (Prentice-Hall, 1989).
29. J. F. Cai, S. Osher, and Z. Shen, "Split Bregman methods and frame based image restoration," Multiscale Model. Simul. **8**, 337–369 (2009).
30. A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," SIAM J. Imaging Sci. **2**, 183–202 (2009).
31. 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. **19**, 2345–2356 (2010).
32. M. Figueiredo, J. Bioucas-Dias, and R. Nowak, "Majorization-minimization algorithms for wavelet based image restoration," IEEE Trans. Image Process. **16**, 2980–2991 (2007).
33. Th. Kreis, *Handbook of Holographic Interferometry: Optical and Digital Methods* (Wiley-VCH, 2005).
34. "Multi-plane phase retrieval," http://www.cs.tut.fi/~lasip/DDT/.