

Augmented Lagrangian complex wave field reconstruction from intensity-only data

Artem Migukin*, Vladimir Katkovnik, and Jaakko Astola

Signal Processing Institute, University of Technology of Tampere,

P. O. Box 553, Tampere, Finland.

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

A new recursive augmented Lagrangian (AL) algorithm is presented for reconstruction of a 3D wave field for intensity-only measurements obtained from two or more sensor planes parallel to the object plane. This reconstruction is framed as a maximum likelihood constrained nonlinear optimization problem for Gaussian additive noise observations. A contribution of this paper concerns a development of a novel recursive algorithm and demonstration that this algorithm enables a better accuracy and better imaging comparing with the successive iterative method originated in [Opt. Lett. **30**, 833 (2005)]. © 2013 Optical Society of America

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

1. Introduction

Let $u_0(x)$ and $u_r(x)$, $r = \overline{1, K}$, denote complex-valued wave field distributions at the object and observation (sensor) planes respectively, given in lateral coordinates $x \in R^2$. The index r corresponds to a distance $z_r = z_1 + (r - 1) \cdot \Delta_z$ between the parallel object and the r -th observation planes, Δ_z is a distance between the observation planes, and K is a number of the

observation planes. It is assumed that the wave field distributions at the object and sensor planes are pixelated, i.e. they are pixel-wise 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: $u_0(x) \rightarrow u_0[k]$, $u_r(x) \rightarrow u_r[k]$ with an $2D$ integer argument k .

Discrete intensity observations are given in the form

$$o_r[k] = |u_r[k]|^2 + \varepsilon_r[k], \quad r = \overline{1, K}, \quad (1)$$

where the wave field intensity (power) is measured with an additive random errors $\varepsilon_r[k]$. In this work we assume that the noise is zero-mean Gaussian, with the standard deviation σ_r for the r -th plane, $\varepsilon_r[k] \sim \mathcal{N}(0, \sigma_r^2)$.

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

In what follows we use a vector-matrix notation with complex-valued images (of the size $N \times M$) given as \mathbb{C}^n vectors. Recall that \mathbb{C}^n means a space of complex-valued vector variables of the length $n = N \cdot M$ obtained from usual matrix representations by concatenating the columns of these matrices. Bold lower case characters are used for these vectors. A forward wave field prop-

agation from a diffraction (object) plane with a complex-valued distribution \mathbf{u}_0 gives a complex-valued distribution \mathbf{u}_r at the r -th image (sensor) plane as

$$\mathbf{u}_r = \mathbf{A}_r \cdot \mathbf{u}_0, \quad (2)$$

where \mathbf{A}_r is a forward propagation operator (complex-valued $n \times n$ matrix, $\mathbf{A}_r \in \mathbb{C}^{n \times n}$) from the object to the r -th plane.

We consider a coherent light scenario with a paraxial wave field propagation model based on the Rayleigh-Sommerfield integral. The operator \mathbf{A}_r in (2) is specified by discretization of this integral. Depending on used discretization this discrete forward propagation model can be: convolutional single or double size [1], angular spectrum decomposition (ASD) [2], or a recent discrete diffraction transform (DDT) given in the matrix (M-DDT) [3] or frequency domain (F-DDT) forms [4]. 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 .

In the vector-matrix notation (2) the equation (1) takes the form

$$\mathbf{o}_r = |\mathbf{u}_r|^2 + \boldsymbol{\varepsilon}_r, \quad r = \overline{1, K}, \quad (3)$$

where the modulus $|\cdot|$ and square $|\cdot|^2$ are the point-wise operations applied

to the elements of the corresponding vectors.

The phase retrieval as it is considered in this paper is a reconstruction of \mathbf{u}_0 from the observations (3). Then the wave field can be reconstructed both for the sensor planes as well as for planes with any axial positions.

In 1972 Gerchberg and Saxton [5] proposed a simple and successful iterative algorithm for solving phase retrieval problems in two dimensions. In a series of the papers which were generalized and unified in [6] Fienup introduced a broad framework for the iterative phase retrieval with few novel algorithms: error-reduction and gradient style. The convergence and the convergence rates of the algorithms are considered in [6].

In a number of publications theoretical aspects of the phase retrieval are studied. Here are the accurate mathematical formulation of the problem, existence and uniqueness of the solution, convergence and interpretation of the known algorithms in terms of mathematical analysis (e.g. [7], [8], [9], where further references can be seen).

New connections between these phase retrieval schemes and classical convex optimization methods are established in [10]. It is shown that these algorithms can be identified as a nonconvex alternating projection algorithm. In

particular, the recursive Gerchberg-Saxton-Fienup algorithms can be interpreted as the recursive projections both at the image and object planes. This interpretation highlights the nature of these algorithms as well as gives an opportunity to design the novel ones with flexible use of extra prior information on measurements and reconstructed distributions [11].

Despite a generosity of the Gerchberg-Saxton-Fienup algorithms the phase reconstruction from multiple plane observations is a special problem and requires a special consideration.

Here we want to discuss few methods especially developed for this problem. A variational approach developed in [12], [13] is based on minimization of the relative entropy (Kullback–Leibler divergence) criterion. It assumes that the distributions of the object plane $|\mathbf{u}_0|^2$ and image planes $|\mathbf{u}_r|^2$ are random and the criterion has a form

$$KL = \sum_r \int_{R^2} |\mathbf{u}_r(x)|^2 \lg \frac{|\mathbf{u}_r(x)|^2}{|\mathbf{u}_0(x)|^2} dx, \quad (4)$$

The estimate is found as a solution of the constrained optimization problem

$$\hat{\mathbf{u}}_0(x) = \min_{\mathbf{u}_0} KL, \quad (5)$$

provided $\int_{R^2} |\mathbf{u}_r(x)|^2 dx = \int_{R^2} |\mathbf{u}_0(x)|^2 dx, r = \overline{1, K}$.

The recursive algorithm derived in [12] generates the recursive estimates $\hat{\mathbf{u}}_0^{(p)}(x)$. In the vector-matrix notation this algorithm is of the form

$$\begin{aligned}\hat{\mathbf{u}}_0^{(p+1)} &= \frac{1}{L} \sum_r \mathbf{A}_r^H \cdot \left[\frac{\mathbf{o}_r}{|\hat{\mathbf{u}}_r^{(p)}|^2} \circ \hat{\mathbf{u}}_r^{(p)} \right], \\ \hat{\mathbf{u}}_r^{(p)} &= \mathbf{A}_r \cdot \hat{\mathbf{u}}_0^{(p)}, p = 0, 1, \dots\end{aligned}\quad (6)$$

Here \mathbf{A}_r is a matrix representation of ASD, and $\mathbf{A}_r^H = (\mathbf{A}_r^*)^T$ is the Hermitian transpose of \mathbf{A}_r corresponding to the ASD backward propagation. The observations \mathbf{o}_r are noiseless, $\mathbf{o}_r = |\mathbf{A}_r \hat{\mathbf{u}}_0|^2$.

In (6) the operations in the square brackets are Hadamard elementwise, i.e. $\frac{\mathbf{o}_r}{|\hat{\mathbf{u}}_r^{(p)}|^2}$ is a vector obtained by the elementwise division of the vectors \mathbf{o}_r and $|\hat{\mathbf{u}}_r^{(p)}|^2$. Multiplication of the vectors $\frac{\mathbf{o}_r}{|\hat{\mathbf{u}}_r^{(p)}|^2}$ and $\hat{\mathbf{u}}_r^{(p)}$ is also elementwise. Thus, $\left[\frac{\mathbf{o}_r}{|\hat{\mathbf{u}}_r^{(p)}|^2} \circ \hat{\mathbf{u}}_r^{(p)} \right]$ is a vector.

Comparison with the conventional Missel's algorithm [14] (similar to Gerchberg-Saxton-Fienup technique) is discussed in [12]. In our notation this algorithm is of the form

$$\begin{aligned}\hat{\mathbf{u}}_0^{(p+1)} &= \frac{1}{L} \sum_{r=1}^K \mathbf{A}_r^H \cdot \left[\frac{\sqrt{\mathbf{o}_r}}{|\hat{\mathbf{u}}_r^{(p)}|} \circ \hat{\mathbf{u}}_r^{(p)} \right], \\ \hat{\mathbf{u}}_r^{(p)} &= \mathbf{A}_r \cdot \hat{\mathbf{u}}_0^{(p)}, p = 0, 1, \dots\end{aligned}\quad (7)$$

Here, multiplication (again elementwise as it is in (6)) of $\hat{\mathbf{u}}_r^{(p)}$ by $\frac{\sqrt{\mathbf{o}_r}}{|\hat{\mathbf{u}}_r^{(p)}|}$ means a replacement of the module of $\hat{\mathbf{u}}_r^{(p)}$ by $\sqrt{\mathbf{o}_r}$, typically as it is in the

Gerchberg-Saxton-Fienup algorithms. After that the matrix \mathbf{A}_r^H backward propagates this estimate to the object plane. The final estimate $\hat{\mathbf{u}}_0^{(p+1)}$ in (7) is a sample mean of these backward estimates calculated for sensor planes.

The algorithm (6) is different from (7) by the weights for $\hat{\mathbf{u}}_r^{(p)}$. Instead of the ratio of the modules in (7) the ratio of the powers (intensities) is appeared in (6).

An algorithm of the structure similar to (7) is derived in [15] from different speculations. First, a linear least square estimate is obtained for $\hat{\mathbf{u}}_0$ assuming that complex-valued observations are available for the sensor plane. After that the modules of these artificial sensor data are replaced by the square roots of the measurements as it is in (7). This algorithm has a form

$$\begin{aligned} \hat{\mathbf{u}}_0^{(p+1)}(x) &= \left(\sum_{r=1}^K \frac{1}{\sigma_r^2} \mathbf{A}_r^H \mathbf{A}_r + \mu \cdot \mathbf{I}_{n \times n} \right)^{-1} \times \\ &\times \sum_{r=1}^K \frac{1}{\sigma_r^2} \mathbf{A}_r^H \cdot \left[\frac{\sqrt{\mathbf{O}_r}}{|\hat{\mathbf{u}}_r^{(p)}|} \circ \hat{\mathbf{u}}_r^{(p)} \right], \\ \hat{\mathbf{u}}_r^{(p)} &= \mathbf{A}_r \cdot \hat{\mathbf{u}}_0^{(p)}, \quad p = 0, 1, \dots \end{aligned} \quad (8)$$

where $\mu > 0$ is a regularization parameter and σ_r^2 is the variance of the noise at the r -th plane.

For ASD $\mathbf{A}_r^H \mathbf{A}_r = \mathbf{I}_{n \times n}$. Let $\mu = 0$, $\sigma_r^2 = \sigma^2$ then (8) coincides with (7).

However, the algorithm (8) is applicable for any forward propagation model

which can be different from ASD, in particular, for the *DDT* models, which do not have an accurate inverse. Recall that ASD used (6) and (7) has an accurate inverse in the matrix form given as \mathbf{A}_r^H .

Recently, a novel circular recursive phase retrieval algorithm for multiple plane observations is proposed in [16], [17]. It is known as the single-beam multiple-intensity phase reconstruction (SBMIR) algorithm. Let the wave field at the r -th and $(r + 1)$ -th planes are 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 in the corresponding planes. Then the forward-forward algorithm from [18] can be written in the form

1. Repeat for $p = 0, 1, \dots$ (9)

2. For $r = \overline{1, K - 1}$

$$\hat{\mathbf{u}}_r^{(p)} = \mathbf{A}_{r+1,r} \cdot \left[\frac{\sqrt{\mathbf{o}_r}}{|\hat{\mathbf{u}}_r^{(p)}|} \circ \hat{\mathbf{u}}_r^{(p)} \right]$$

3. For $r = K$

$$\hat{\mathbf{u}}_1^{(p+1)} = \mathbf{A}_{1,K} \cdot \left[\frac{\sqrt{\mathbf{o}_K}}{|\hat{\mathbf{u}}_K^{(p)}|} \circ \hat{\mathbf{u}}_K^{(p)} \right]$$

4. End on r

5. End on p

Similar to the formulas (6)-(8) 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 $r = \overline{1, K}$. In step 3 of the algorithm (9) the wave field at the first image plane is estimated from the estimate for the last $K - th$ plane, where $\mathbf{A}_{1,K}$ is a backward propagation operator from the $K - th$ sensor plane to the first one.

The algorithm (9) updates periodically the phase distributions at the observation planes. It makes a serious difference versus the algorithms (6)-(8), where the iterations concern the object distribution only.

2. Proposed variational approach

In this article, we follow the maximum likelihood (ML) style approach.

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

$$J = \sum_r \frac{1}{2\sigma_r^2} \|\mathbf{o}_r - |\mathbf{u}_r|^2\|_2^2 + \mu \cdot pen(\mathbf{u}_0), \quad (10)$$

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

The first summand in (10) is the quadratic fidelity term obtained as the main term of the minus logarithm of the Gaussian likelihood function corresponding to the model (3), and the second term is the penalty (regularization)

including prior information (simpler the prior) on the object distribution \mathbf{u}_0 to be reconstructed.

Different metrics for the fidelity terms concerning the fitting of observations are used for the phase retrieval. The most popular is a comparison of the observed and predicted modules of the wave field distributions. For (10) it means $\sum_r \|\sqrt{\mathbf{o}_r} - |\mathbf{u}_r|\|_2^2$.

A more general construction of the form $\sum_r \|\mathbf{o}_r^{\gamma/2} - |\mathbf{u}_r|^\gamma\|_2^2$, where γ is a parameter, is discussed in [19] and [20]. Some mathematical aspects of the difference between this criteria with $\gamma = 1$ and $\gamma = 2$ are discussed in [8]. This discussion can be summarized as follows. For $\gamma = 1$, the criterion is not differentiable. It makes the mathematical analysis more difficult and potentially results in a lower convergence rate. For $\gamma > 1$, the criterion becomes differentiable and more convenient for mathematical analysis. However, the experiments show that the accuracy for $\gamma = 1$ is better than for $\gamma = 2$. As a compromise to $\gamma = 1$ the criterion $\sum_r \|\mathbf{o}_r^{1/2} - |\mathbf{u}_r|^2 / \sqrt{|\mathbf{u}_r|^2 + \delta}\|_2^2$ is considered as a differentiable function, where $\delta > 0$ is small.

More complex criteria are used for superresolution phase retrieval with multiple subarrays of sensors [20], [21]. The corresponding criteria usually

have a form of sums squared residuals for the subarrays.

In the criterion (10) the choice is done in favor $\gamma = 2$ as it is dictated by the statistical speculations. The squared residuals $(\mathbf{o}_r - |\mathbf{u}_r|^2)$ in (10) are appeared due to the ML technique provided that the observation noise is Gaussian, and the sum of the residuals appeared due to assumption that the noises are independent.

Using the criterion (10) we formulate the wave field reconstruction as the following constrained optimization

$$\begin{aligned} \hat{\mathbf{u}}_0 = & \tag{11} \\ \arg \min_{\mathbf{u}_0} \sum_r \frac{1}{2\sigma_r^2} \|\mathbf{o}_r - |\mathbf{u}_r|^2\|_2^2 + \mu \cdot pen(\mathbf{u}_0) \\ \text{subject to } & \mathbf{u}_r = \mathbf{A}_r \mathbf{u}_0. \end{aligned}$$

The parameter μ in (11) defines a balance between the accuracy of the observation fitting and the prior given by the penalty $pen(\mathbf{u}_0)$. 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 non-smooth $\hat{\mathbf{u}}_0$. If $\mu > 0$ and large then the noise effects are well suppressed but the solution $\hat{\mathbf{u}}_0$ can be oversmoothed with important features lost.

A proper selection of μ known as a regularization parameter is an important

point of the variation formulation in inverse imaging.

Numerous forms for the penalty $pen(\cdot)$ have been used in literature on digital image processing derived from speculations varying from probabilistic modeling of image distribution priors to heuristic constructions (e.g. [22], [23]).

In this paper we use a simple and sometimes quite efficient quadratic Tikhonov's penalty [24] given in the form

$$pen(\mathbf{u}_0) = \|\mathbf{u}_0\|_2^2. \quad (12)$$

3. Augmented Lagrangian (AL)

Inserting $\mathbf{u}_r = \mathbf{A}_r \mathbf{u}_0$ in the criterion (11) we arrive from the constrained optimization to the unconstrained one $\hat{\mathbf{u}}_0 = \arg \min_{\mathbf{u}_0} \sum_r \frac{1}{2\sigma_r^2} \|\mathbf{o}_r - \mathbf{A}_r \mathbf{u}_0\|_2^2 + \mu \cdot pen(\mathbf{u}_0)$. At first glance it looks simpler than the constrained one (11). Various gradient, Newton and Gauss-Newton recursive techniques can be applied to find this unconstrained solution. The main drawback of this approach not only the calculation complexity of the gradient and second derivative matrices but a typically low convergence rate of the corresponding recursive procedures. It is recognized that in general the constrained optimization can result in more efficient algorithms.

The augmented Lagrangian (AL) method, introduced independently by Hestenes [25] and Powell [26] is now classical for the minimization of functionals in presence of linear equality constraints.

AL corresponding to the problem (11) with complex-valued variables is of the form

$$\begin{aligned}
 L(\mathbf{u}_0, \{\mathbf{u}_r\}, \{\boldsymbol{\Lambda}_r\}) = & \tag{13} \\
 & \sum_{r=1}^K \frac{1}{\sigma_r^2} \left[\frac{1}{2} \|\mathbf{o}_r - |\mathbf{u}_r|^2\|^2 + \frac{1}{\gamma_r} \|\mathbf{u}_r - \mathbf{A}_r \cdot \mathbf{u}_0\|^2 + \right. \\
 & \left. + \frac{2}{\gamma_r} \operatorname{Re}\{\boldsymbol{\Lambda}_r^H (\mathbf{u}_r - \mathbf{A}_r \cdot \mathbf{u}_0)\} \right] + \mu \|\mathbf{u}_0\|_2^2,
 \end{aligned}$$

where $\boldsymbol{\Lambda}_r \in \mathbb{C}^n$ and $\boldsymbol{\Lambda}_r^H = (\boldsymbol{\Lambda}_r^*)^T$.

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 $\{\boldsymbol{\Lambda}_r\}$. The parameters γ_r are positive.

In AL both the linear and quadratic terms corresponding to the linear constraints $\mathbf{u}_r - \mathbf{A}_r \cdot \mathbf{u}_0 = 0$ are used. If we keep only the quadratic terms the augmented Lagrangian becomes the penalty criterion, which can be used assuming that the penalty coefficient $1/\gamma_1$ 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 augmented Lagrangian becomes the standard

Lagrangian. However, the saddle-point of this standard Lagrangian is unstable. It results in the problems of numerical solutions. The stability of the saddle-point of the augmented Lagrangian is one of the principal advantages of this criterion.

The successive steps of augmented Lagrangian optimization are as follows

$$(\mathbf{u}_{0,t+1}, \{\mathbf{u}_{r,t+1}\}) \in \arg \min_{\mathbf{u}_0, \{\mathbf{u}_r\}} L(\mathbf{u}_0, \{\mathbf{u}_r\}, \{\Lambda_{r,t}\}), \quad (14)$$

$$\Lambda_{r,t+1} = \Lambda_{r,t} + \alpha_r \cdot (\mathbf{u}_{r,t} - \mathbf{A}_r \cdot \mathbf{u}_{0,t}), \quad (15)$$

where minimization on $\mathbf{u}_0, \{\mathbf{u}_r\}$ (14) is produced for fixed $\{\Lambda_{r,t}\}$. The steps on the Lagrangian multipliers (15) are produced in the gradient direction (for gradient calculated as $\nabla_{\Lambda_r} L$) with the step size α_r .

This leads to the alternating minimization on $\mathbf{u}_0, \{\mathbf{u}_r\}$ and to the algorithm of the form

For $t = 0, 1, \dots$

$$\mathbf{u}_{r,t+1} \in \arg \min_{\{\mathbf{u}_r\}} L(\mathbf{u}_{0,t}, \{\mathbf{u}_{r,t}\}, \{\Lambda_{r,t}\}), \quad (16)$$

$$\mathbf{u}_{0,t+1} \in \arg \min_{\mathbf{u}_0} L(\mathbf{u}_{0,t}, \{\mathbf{u}_{r,t+1}\}, \{\Lambda_{r,t}\}), \quad (17)$$

$$\Lambda_{r,t+1}^* = \Lambda_{r,t}^* + \alpha_r \cdot (\mathbf{u}_{r,t+1} - \mathbf{A}_r \cdot \mathbf{u}_{0,t+1}), \quad (18)$$

End on t

This type of the algorithms recently become a subject of an intensive development and study, in particular with application to image processing and focused on l_1 -norm prior. These algorithms sharing many common ideas and features are appeared under different names such as split Bregman iterations [27], iterative shrinkage-thresholding algorithms [28], alternating direction method of multipliers [29], majorization-minimization algorithms [30]. In this paper we are restricted to the augmented Lagrangian techniques pragmatically using it for derivation of the proposed algorithms.

Let us perform calculations defining our algorithm.

3.1. Optimization on \mathbf{u}_0

Minimization on a complex-valued $\mathbf{u}_0 = \text{Re}\{\mathbf{u}_0\} + i \text{Im}\{\mathbf{u}_0\}$ ($\mathbf{u}_0 = \mathbf{a}_0 + i\mathbf{b}_0$) 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 $\partial L/\partial \mathbf{a}_0 = 0$ and $\partial L/\partial \mathbf{b}_0 = 0$. However, it is more convenient, and it results in more compact equations 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 $\partial L/\partial \mathbf{u}_0^* = 0$ or equivalently $\partial L/\partial \mathbf{u}_0 = 0$. In these derivative calculations the variables

\mathbf{u}_0 and \mathbf{u}_0^* are treated as independent. Note also that differentiation by vector gives a vector of derivatives, $\partial L/\partial \mathbf{u}_0 = [\partial L/\partial \mathbf{u}_0[1], \dots, \partial L/\partial \mathbf{u}_0[n]]^T$, provided that $\mathbf{u}_0 = [\mathbf{u}_0[1], \dots, \mathbf{u}_0[n]]^T$. 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 $\partial \|\mathbf{u}_0\|_2^2/\partial \mathbf{u}_0 = \mathbf{u}_0^*$.

Let us rewrite the criterion (13) into the form

$$\begin{aligned}
L(\mathbf{u}_0, \{\mathbf{u}_r\}, \{\mathbf{\Lambda}_r\}) = & \tag{19} \\
& \sum_{r=1}^K \frac{1}{\sigma_r^2} \left\{ \frac{1}{2} \|\mathbf{o}_r - |\mathbf{u}_r|^2\|^2 + \frac{1}{\gamma_r} \|\mathbf{u}_r - \mathbf{A}_r \cdot \mathbf{u}_0\|^2 + \right. \\
& \left. + \frac{1}{\gamma_r} [\mathbf{\Lambda}_r^H (\mathbf{u}_r - \mathbf{A}_r \cdot \mathbf{u}_0) + (\mathbf{u}_r - \mathbf{A}_r \cdot \mathbf{u}_0)^H \mathbf{\Lambda}_r] \right\} + \\
& + \mu \|\mathbf{u}_0\|_2^2.
\end{aligned}$$

We use the minimum condition for (19) in the form $\nabla_{\mathbf{u}_0^*} L = 0$. It gives

$$\sum_{r=1}^K \frac{1}{\gamma_r \sigma_r^2} \mathbf{A}_r^H (\mathbf{A}_r \mathbf{u}_0 - \mathbf{u}_r) - \frac{1}{\gamma_r \sigma_r^2} \mathbf{A}_r^H \mathbf{\Lambda}_r + \mu \mathbf{u}_0 = 0.$$

and further

$$\begin{aligned}
\hat{\mathbf{u}}_0 = & \left(\sum_{r=1}^K \frac{1}{\gamma_r \sigma_r^2} \mathbf{A}_r^H \mathbf{A}_r + \mu \cdot \mathbf{I}_{n \times n} \right)^{-1} \times \\
& \times \sum_{r=1}^K \frac{1}{\gamma_r \sigma_r^2} \mathbf{A}_r^H (\mathbf{u}_r + \mathbf{\Lambda}_r).
\end{aligned} \tag{20}$$

For the circular convolution the matrix \mathbf{A} is a structured Toeplitz matrix and solution of (20) can be implemented using \mathcal{FFT} .

3.2. Optimization on \mathbf{u}_r

The Lagrangian (19) 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 $\nabla_{\mathbf{u}_r^*[k]}L = 0$ gives the minimum condition for $\mathbf{u}_r[k]$ as

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{u}_r^*[k]} &= \frac{1}{\sigma_r^2} (|\mathbf{u}_r[k]|^2 - \mathbf{o}_r[k]) \cdot \mathbf{u}_r[k] + \\ &+ \frac{1}{\gamma_r \sigma_r^2} (\mathbf{u}_r[k] - (\mathbf{A}_r \cdot \mathbf{u}_0)[k] + \Lambda_r[k]) = 0. \end{aligned} \quad (21)$$

It follows

$$\mathbf{u}_r[k] = \frac{(\mathbf{A}_r \cdot \mathbf{u}_0)[k] - \Lambda_r[k]}{\gamma_k (|\mathbf{u}_r[k]|^2 - \mathbf{o}_r[k]) + 1}. \quad (22)$$

Then for the module

$$|\mathbf{u}_r[k]| = \frac{|(\mathbf{A}_r \cdot \mathbf{u}_0)[k] - \Lambda_r[k]|}{|\gamma_k (|\mathbf{u}_r[k]|^2 - \mathbf{o}_r[k]) + 1|}. \quad (23)$$

The last equation is cubic with respect to $|\mathbf{u}_r[k]|$. It can have a single or three real solutions. We are looking for a nonnegative $|\mathbf{u}_r[k]|$. Let us denote the desirable solution of (23) as $|\tilde{\mathbf{u}}_r[k]|$. Then, the corresponding complex-valued $\mathbf{u}_r[k]$ is calculated from (22) as

$$\mathbf{u}_r[k] = \frac{(\mathbf{A}_r \cdot \mathbf{u}_0)[k] - \Lambda_r[k]}{\gamma_k (|\tilde{\mathbf{u}}_r[k]|^2 - \mathbf{o}_r[k]) + 1}. \quad (24)$$

If the nonnegative solution of (23) is not unique the procedure become more complex as the solutions should be analyzed and we need select a solution

giving a minimum value to the corresponding r -th summand of L : $\frac{1}{2} \|\mathbf{o}_r - |\mathbf{u}_r|^2\|^2 + \frac{1}{\gamma_r} \|\mathbf{u}_r - \mathbf{A}_r \cdot \mathbf{u}_0\|^2 + \frac{2}{\gamma_r} \text{Re}\{\mathbf{\Lambda}_r^H (\mathbf{u}_r - \mathbf{A}_r \cdot \mathbf{u}_0)\}$.

Let us denote the nonlinear algorithm giving the solution (24) as

$$\mathbf{u}_r[k] = \mathcal{G}(\mathbf{o}_r[k], \mathbf{u}_r[k], \mathbf{\Lambda}_r[k]), \quad (25)$$

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

3.3. Optimization on $\mathbf{\Lambda}_r$

The gradient $\nabla_{\mathbf{\Lambda}_r^*} L$ is calculated as

$$\nabla_{\mathbf{\Lambda}_r^*} L = \mathbf{u}_r - \mathbf{A}_r \cdot \mathbf{u}_0. \quad (26)$$

3.4. Proposed algorithm

Combing the above equations we arrive to the following AL recursive algorithm:

AL Algorithm (27)

1. Set $p = 0$ (initialization), $\mathbf{u}_{0,0}$, $\mathbf{\Lambda}_{r,0}$,
2. Repeat for $r = \overline{1, K}$, $t = 0, 1, \dots$
3. $\mathbf{u}_{r,t+1/2} = \mathbf{A}_r \cdot \mathbf{u}_{0,t}$,
4. $\mathbf{u}_{r,t+1}[k] = \mathcal{G}(\mathbf{o}_r[k], \mathbf{u}_{r,t+1/2}[k], \mathbf{\Lambda}_{r,t}[k])$,
5. $\mathbf{u}_{0,t+1} = \left(\sum_{r=1}^K \frac{1}{\gamma_r \sigma_r^2} \mathbf{A}_r^H \mathbf{A}_r + \mu \cdot \mathbf{I}_{n \times n} \right)^{-1} \times$
 $\sum_{r=1}^K \frac{1}{\gamma_r \sigma_r^2} \mathbf{A}_r^H (\mathbf{u}_{r,t+1} + \mathbf{\Lambda}_{r,t})$,
6. $\mathbf{\Lambda}_{r,t+1} = \mathbf{\Lambda}_{r,t} + \alpha_r \cdot (\mathbf{u}_{r,t+1} - \mathbf{u}_{r,t+1/2})$,
7. Stop

The initialization concerns the object plane distribution $\mathbf{u}_{0,0}$ and Lagrange vectors $\mathbf{\Lambda}_{r,0}$. Step 3 returns the forward propagation of the object wave field iteration $\mathbf{u}_{0,t}$ to the sensor planes, and step 4 is used to agree these estimates with the observations $\mathbf{o}_r[k]$ according to the estimate obtained using (25). Step 5 gives the update for the object wave field calculated from the image

estimates $\mathbf{u}_{r,t+1}$ and the Lagrange multipliers $\Lambda_{r,t}$. This step is similar to the algorithm (8). Step 6 returns updated values of the Lagrange multipliers.

In conclusion of this section we wish to note that a similar AL algorithm is developed in [32] for 4f optical setup with phase-random optical masks in the Fourier focal plane of the first lens. This algorithm is used for the object wave field reconstruction from multiple intensity observations different by optical masks with independent phase modulation. The algorithm in [32] is derived in the frequency Fourier domain what is essentially different from the derivation given above.

4. Simulation experiments

The main goal of these experiments is to analyze the accuracy of wave field reconstruction for various parameters $\{\gamma_r, \alpha_r, \mu\}$ of the AL algorithm and to study the algorithm performance for different multi-plane phase retrieval scenarios. First of all we are looking for suitable parameters of the AL algorithm, which can be used for different experiments and for both noisy and noiseless observations (3). The wave field reconstruction accuracy is characterized via the root-mean-square error (*RMSE*) criterion.

We calculate the *RMSE* values for the amplitude and phase of the wave

field distribution separately, moreover $RMSE$ for the phase is calculated for the unbiased phase estimate $\hat{\phi}_0 - \text{mean}(\hat{\phi}_0)$.

The influence of quantization of the observations is out of the scope in this work, and we assume that a double precision data $\{\mathbf{o}_r\}_{r=1}^K$ are given. The number of iterations of the algorithm is fixed and equal to 100. A comparison of the developed algorithm is produced versus the SBMIR algorithm [17], [18].

The numerical experiments are performed for the amplitude and phase modulation of the object distribution $u_0[k] = |u_0[k]| \cdot \exp(j \cdot \phi_0[k])$. For the amplitude modulation (AM) $\phi_0[k] \equiv 0$, and for the phase modulation (PM) $|u_0[k]| \equiv 1$. We assume that $|u_0[k]| = w[k]$ for AM and $\phi_0[k] = \pi(w[k] - \frac{1}{2})$ for PM , where the test-image is scaled in such a way that $0 < w[k] \leq 1$. We present results of numerical experiments for the following square $N \times N$ test-images: gray-scale *lena* (256×256), binary *logo* (256×256) and *chessboard* (128×128), and the geometrical smooth *Mexican Hat* (200×200)

$$u_0(x) = -\frac{4}{3\sqrt{\pi}} \cdot (\|x\|^2 - 2) \cdot \exp(-\frac{1}{2}\|x\|^2), \quad (28)$$

defined on the 2D discrete grid $-6 \leq (x_1, x_2) \leq 6$ with the step size 0.025.

The last test-image is normalized in order to make it nonnegative.

Pixelated object and sensor plane distributions have square pixels $\Delta \times \Delta$ of

the equal size (we consider $\Delta = 6.7\mu m$) and 100% fill factors. The wavelength is $\lambda = 532nm$, which corresponding to a Nd:YAG green laser. The "in-focus" distance is calculated as $z_f = N \cdot \Delta^2/\lambda$ (see [3]). The distance between the object plane and the first parallel sensor plane z_1 is expressed through this "in-focus" distance as $z_1 = d \cdot z_f$. In our simulations the parameter d is from the interval $[0.5, 3]$. In most experiments $d = 2$. A number of measurement planes K varies in the interval $[2, 15]$. Mainly $K = 5$.

Our experiments are made for different distances Δ_z between the observation planes. We analyze the best choice of Δ_z for different settings. If it is not specified $\Delta_z = 2 mm$.

The observations \mathbf{o}_r are generated using the double size *F-DDT* [4] for the forward propagation. One of the motivations in this choice is that this modeling is precise for pixelated sensor and object distributions. The same *F-DDT* forward propagation is used in the AL algorithms for the wave field reconstruction.

The Matlab code used for all simulation experiments is available at <http://www.cs.tut.fi/~lasip/DDT/>. At this website we also provide further results and the original test-images.

4.A. Parameters of the algorithm

In this subsection we present recommended values of the algorithm parameters, which are experimentally found being acceptable for different settings and different test images as well as for different types of modulation at the object plane. For simplicity, we assume that $\alpha_r = \alpha$, $\gamma_r = \gamma$, $\sigma_r^2 = \sigma^2$ for all r . Then, we can rewrite the Step 5 of the algorithm (27) in the following more compact form:

$$\mathbf{u}_{0,t+1} = \left[\sum_{r=1}^K \mathbf{A}_r^H \mathbf{A}_r + \tilde{\mu} \gamma \cdot \mathbf{I}_{n \times n} \right]^{-1} \times \sum_{r=1}^K \mathbf{A}_r^H (\mathbf{u}_{r,t+1} + \mathbf{\Lambda}_{r,t}), \quad (29)$$

where the regularization parameter is changed for $\tilde{\mu} = \mu \cdot \sigma^2$. Thus, we are looking for an acceptable triplet of the AL parameters $\{\gamma, \alpha, \tilde{\mu}\}$. The further discussion about the wave field reconstruction and the choice of the parameters is made for the AL algorithm with the object wave field estimation in the form of (29). Thus, the noise variance σ^2 is imbedded in the parameter $\tilde{\mu}$.

Our numerical experiments have shown that $\mathbf{u}_{0,0} = \frac{1}{2} \cdot \mathbf{1}_{n \times 1}$ is a good initial guess for both the amplitude and the phase modulation of the object distribu-

tion. For the choice of the algorithm parameters and all further experiments this initialization is used.

It is found that an optimal triplet of the parameters $\{\gamma, \alpha, \tilde{\mu}\}$ depends on the wavelength λ only slightly. Thus, we can assume that the choice of these parameters is invariant with respect to λ .

The pixel size Δ is a much more essential parameter. The optimal values of $\tilde{\mu}$ and γ may differ in several times depending on Δ values, especially for *PM*. For smaller Δ the parameter $\tilde{\mu}$ should be larger and the parameter γ should be smaller.

Let us consider a small variation, say about $5 \div 10\%$, in the distance between the observation planes Δ_z . It is found that the difference for the optimal value of γ is insignificant. However, this variation in Δ_z is quite crucial for the regularization parameter $\tilde{\mu}$. Firstly, $\tilde{\mu}$ should not be very small even for noiseless data, we take $\tilde{\mu} \geq 10^{-4}$. Secondly, $\tilde{\mu}$ should be increased for larger Δ_z . Nevertheless, for $z_1 \gg z_f$ and large Δ_z the regularization parameter $\tilde{\mu}$ can be fixed. There are no significant improvements in $\tilde{\mu}$ optimization.

The following values are recommended for the AL triplet $\{\gamma, \alpha, \tilde{\mu}\}$.

The step-size $\alpha > 0$ should be close to 1, but not larger than 1. Its optimal

value varies very little depending on other experiment parameters. $\alpha = 1$ is used in all our simulation experiments.

It is found, that the regularization parameter $\tilde{\mu}$ for *PM* should be in the order larger than that for *AM*. Providing that the level of the noise is small ($\sigma \in [0.01, 0.05]$) it is recommended $\tilde{\mu} = 5 \cdot 10^{-3} / 10^{-2}$ for noisy data and $\tilde{\mu} = 5 \cdot 10^{-4} / 10^{-3}$ if $\sigma = 0$ for *AM/PM*, respectively.

The dependence on the penalty parameter γ is more complex because for various test-images and setups the optimal γ is different. In most our experiments we use the fixed value $\gamma = 10$ for all types of the object distributions.

While in our experiments the parameter are fixed let us discuss the parameter choice with more details.

It is obvious that for particular cases the wave field reconstruction accuracy can be improved by modifying $\{\gamma, \alpha, \tilde{\mu}\}$. As general recommendation we wish to note, that it is reasonable to take larger regularization parameter $\tilde{\mu}$ for larger values of d and/or for a larger number of the measurement planes K .

Note that for different type of test-images the optimal $\tilde{\mu}$ can vary significantly. For instance, for $\sigma = 0$, *AM*, and binary images we recommend to take the regularization parameter in order larger (we used $\tilde{\mu} = 5 \cdot 10^{-3}$ for

logo) than that for natural ones (in our experiments $\tilde{\mu} = 5 \cdot 10^{-4}$ for *lena*).

It is recommended to take γ starting from $\gamma \approx 20/100$ for *AM/PM* and take larger values of γ for larger K and smaller for larger distances z_1 .

It is recommended also to take the step-size α smaller (for instance, select α from the interval $[0.3, 0.9]$) if $z_r < z_f$ or K is small (e.g., $K = 3$).

4.B. Quality of imaging and algorithm performance

In this subsection we consider the AL algorithm performance for both noiseless ($\sigma = 0$) and noisy data ($\sigma = 0.05$) with the fixed parameters $\{\gamma, \alpha, \tilde{\mu}\}$ chosen as it is discussed above.

In Fig. 2 we present a qualitative comparison of the reconstruction imaging for different distances z_1 and different numbers of measurement planes K . These results are given for *AM* with the binary test-image *logo*. The columns of this image corresponds to different K taking values $K = 3, 5, 10$. The rows are given for different distances z_1 varying from $0.5 \cdot z_f$ to $5 \cdot z_f$, $z_f = 21.6mm$ (from the top to bottom row of the table).

It is found that a larger number of K leads to the monotonically better wave field reconstruction. If $z_1 = 0.5 \cdot z_f$ a very good result is obtained already for $K = 5$ (the second image in the top row, $RMSE(|u_0|) = 0.0106$). However the

accuracy and imaging for $K = 10$ are slightly better (the third image in the top row, $RMSE(|u_0|) = 0.0101$). For other rows a larger number of K gives more noticeable improvements in the wave field reconstruction accuracy (see the results in the corresponding table 1) and imaging, while the best quality and the accuracy are obtained for the smallest $d = 0.5$.

Table 1. The quantitative comparison of the amplitude reconstruction, $RMSE(|u_0|)$ for the test, presented in Fig. 2.

$d \backslash K$	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

Let us analyze the reconstruction accuracy in these experiments with *logo* for both the object amplitude $|u_0[k]|$ and phase $\phi_0[k]$. $RMSE(|u_0|)$ and $RMSE(\phi_0)$ are $RMSE$ values for the module and the phase reconstructions, respectively. $E_r(RMSE(|u_0|))$ and $E_r(RMSE(\phi_0))$ are the mean values of these criteria calculated over K sensor planes in order to evaluate the ac-

curacy of the 3D wave field reconstruction at the measurement planes. Fig. 3 illustrates the accuracy of the reconstructions and the convergence rates of the estimates in case of noiseless data ($\sigma = 0$). The solid curves labeled as $E_r(RMSE(|u_r|))$ and $E_r(RMSE(\phi_r))$ correspond to the mean value of $RMSE$ for the amplitude and the phase, respectively, calculated as the mean over the all sensor planes. The vertical intervals around these curves show the standard deviations of the corresponding $RMSEs$ calculated as deviations from the mean over the all sensor planes. We can see the convergence for both the amplitude and phase reconstructions in the means and standard deviations of $RMSEs$. This example shows how the convergence at the sensor planes leads to the convergence at the object plane: compare $RMSE$ for the sensor plane estimates with the corresponding solid curve for $RMSE(|u_0|)$ and the dash curve for $RMSE(\phi_0)$. It can be seen quite a fast decrease (in more than three times) in values of $RMSE(|u_0|)$. The behavior of $RMSE(\phi_0)$ in this case is not so straightforward. It goes up during first 10 iterations and then decreases monotonically but quit slowly.

Let us consider the wave field reconstruction, where the binary test-image *chessboard* is used for the phase of the object distribution and the module

of the object distribution is equal to 1. Fig. 4 illustrates the accuracy of the reconstructions and the convergence rates of the estimates. In this case the data are noisy, $\sigma = 0.05$.

A slow convergence with initial peaks in $RMSE(|u_0|)$ can be noticed for the amplitude reconstruction at the object plane. These peaks are similar to what is observed in Fig. 3 for object phase reconstruction in case of *AM*. Overall the convergence for the phase reconstruction is good but we can note that there is a divergence in the standard deviation calculated for $E_r(RMSE(\phi_r))$.

Optimization of the estimates with respect to the distance Δ_z is a complex problem because it affects the choice of the algorithm parameters and depends on K , z_1 and σ . In Fig. 5 we demonstrate the dependence of the accuracy on Δ_z and existence of the optimal value for this parameter. Here we present the phase reconstruction accuracy for the binary test-image *chessboard*.

The curves are given for noisy data with $\sigma = 0.05$. From smaller Δ_z to the optimal ones the values of $RMSE$ are decreased dramatically, while the values of Δ_z larger than the optimal ones result in quite a slow increasing of $RMSE$. It is seen that the increase of the number of observation planes K , the image size ($N \times N$), and the decrease of the distance z_1 lead to a

proportional decrease of the optimal value of Δ_z .

For larger K the optimum on Δ_z is clearly seen.

Despite the clear decrease of the reconstruction accuracy for larger Δ_z (in Fig. 5 we can see a smooth growth of the *RMSE* curves for larger Δ_z), the visual influence of these parameter values is not essential.

Table 2. The quantitative comparison of the phase reconstruction (via $RMSE(\phi_0)$) for the test, presented in Fig. 6.

$K \setminus d$	0.5	1	2
3	0.1111	0.2819	0.5874
5	0.0424	0.1862	0.3603
10	0.0307	0.1361	0.2882

In Fig. 6 the visual imaging for the considered experiment is shown. The corresponding *RMSE* values are presented in Table 2. The images in Fig. 6 vary in the number of observation planes K and the distance z_1 . The images from the top to the bottom correspond to $K = 3, 5, 10$. The columns from left to right correspond to increasing $z_1 = z_f \cdot d$, $d = 0.5, 1, 2$. For different K we use approximately optimal Δ_z calculated as $\Delta_z = 0.4 \cdot z_f$ for $K = 3$, $\Delta_z = 0.3 \cdot z_f$ for $K = 5$ and $\Delta_z = 0.15 \cdot z_f$ for $K = 10$. These approximations

are obtained from Fig. 5. It is noticed that even we take an approximately optimal value of Δ_z it will not improve the imaging significantly because the blurring effects depend mainly on d .

In general, the distances between the observation planes can be chosen as irregular. However, our experiments show that even after optimization of the algorithm parameters the accuracy improvement is not essential approximately 10 ÷ 20% in *RMSE* values.

A larger number of the observation planes K results in monotonically better convergence rate and better accuracy. The relative improvement is especially valuable for small K and not so essential for larger K . For instance, two additional planes from $K = 3$ to $K = 5$ improves *RMSE* approximately by 70% and 30% for the reconstruction of the amplitude and phase, respectively. Five more planes (from $K = 5$ to $K = 10$) results in a smaller improvement by less than 40% and 25% for the reconstruction of the amplitude and phase, respectively. A further increasing of K does not give essential effects.

4.C. Computational complexity of AL algorithm

In Table 3 we present the average calculation time (in seconds) for 100 iteration of the AL algorithm depending on N (the image size in pixels is $n = N^2$)

and the number of the measurement planes K .

Table 3. The computational time (in sec) for 100 iterations of the AL algorithm.

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

The averaged computational times required for calculation of $\{\mathbf{A}_r\}_{r=1}^K$ are shown in Table 4. Note, that these matrices are calculated only ones for reconstruction and can be calculated in advance and saved for use.

These computational times are obtained by Monte-Carlo simulation with averaging over 50 experiments with 100 iterations both for AM and PM . The computer used for experiments is Intel Core 2Duo E8400 @ 3GHz, RAM: 4GB, Windows Xp SP3, Matlab 7.9.0 (R2009b). Memory requirements depend mainly on the size of the test image and the number of observation planes K . Then required memory is calculated as $N^2 \times 8 \times K \times 2 \times 2 \times 2$

Table 4. The computational time (in sec) for calculation of K wave field propagation operators with double size $F - DDT$.

$K \setminus N$	128	256	512
2	0.41	0.99	1.96
3	0.61	1.27	2.87
5	0.95	1.99	4.55
7	1.24	2.65	6.16
10	1.68	3.65	8.56
15	2.4	5.34	12.69

Bytes for the forward wave field propagation according to (2). Here 8 Bytes stands for double-precision floating point calculations, the multiplier K appears because we have K observation planes, the first factor 2 appears because we use double size F-DDT propagation model for accurate and fast calculations in the frequency domain, the second 2 - because all variables are complex-valued, and the last 2 is a number of matrices in the product.

4.D. Comparison vs SBMIR algorithm

The comparison of the proposed algorithm is produced versus the SBMIR algorithm [16]. The main procedure of this algorithm is given by (9).

The results obtained by the AL and SBMIR algorithms for *AM* object distribution ($\sigma = 0$) are shown in Fig. 7 for reconstruction of the module $|u_0|$ and the phase ϕ_0 at the object plane. The first column is obtained by AL, $\tilde{\mu} = 5 \cdot 10^{-4}$: Fig. 7(a) is the amplitude reconstruction, $RMSE(|u_0|) = 0.041$, and Fig. 7(c) is for the phase reconstruction, $RMSE(\phi_0) = 0.091$. The second column images are obtained by SBMIR. Fig. 7(b) is for the amplitude reconstruction, $RMSE(|u_0|) = 0.08$, and Fig. 7(d) for the phase reconstruction, $RMSE(\phi_0) = 0.28$. The phase reconstructions are shown after the phase unwrapping produced using the ZpM algorithm [31]. The accuracy of the AL algorithm for both the amplitude and phase reconstructions is approximately twice better than that for the SBMIR algorithm. Visually, the advantage of the AL algorithm is well seen for the amplitude reconstruction. 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 comparison of the convergence rates of the AL and SBMIR algorithms for this experiment is presented in Fig. 8.

In Fig. 9 we compare the wave field reconstruction for the phase modulated binary test-image. We consider the case of noisy data, $\sigma = 0.05$, for $K = 5$.

The first column of Fig. 9 is obtained by AL, $\tilde{\mu} = 0.01$: Fig. 9(a) shows the amplitude reconstruction, $RMSE(|u_0|) = 0.23$, and Fig. 9(c) phase reconstruction, $RMSE(\phi_0) = 0.26$. The second column in Fig. 9 is obtained by SBMIR: Fig. 9(b) is for the amplitude reconstruction, $RMSE(|u_0|) = 0.35$, and Fig. 9(d) is for the phase reconstruction, $RMSE(\phi_0) = 0.58$.

The AL algorithm demonstrates clear and sharp phase imaging, while the SBMIR phase reconstruction is blurred and significantly destroyed. The $RMSE$ values for the amplitude $|u_0|$ and phase ϕ_0 are quite high because of the large errors on the borders of the blocks of the *chessboard*. Nevertheless, the reconstruction accuracy both for the amplitude and phase is much better for the AL algorithm. We compare the convergence rates of the AL and SBMIR algorithms for the phase reconstruction in Fig. 10. At the first stage SBMIR demonstrates a better accuracy. However, further its convergence becomes very slow and after 100 iterations AL shows much better results.

In contrast to the discontinuous and non-smooth distributions (*lena*, *chessboard* and *logo* test images, considered above) further we consider the experiment produced for the continuous smooth test-image "Mexican Hat". This image is used for *PM* experiments. In Fig. 11 we show the true phase and

reconstructions, obtained by the AL and SBMIR algorithms. These phase reconstructions are shown after phase unwrapping produced using the ZpM algorithm [31]. The advantage of the AL algorithm is obvious.

The convergence rates of the AL and SBMIR algorithms for this *PM* experiment are illustrated in Fig. 12 with a number of iterations up to 1000.

The AL algorithm converges much faster than SBMIR and gives essentially better result. The almost complete bell and the hollow of *Mexican Hat* are reconstructed by AL (see Fig. 11(b)). However, there are significant errors on the borders in the AL and SBMIR reconstructions.

5. Conclusion

In this work we present a novel variational formulation of the phase retrieval problem. Being the maximum likelihood style this setting takes into consideration the noise distribution. Based on the augmented Lagrangian technique we developed the recursive constrain optimization algorithm for the amplitude and phase reconstruction assuming that the object distribution is complex-valued.

Multiple numerical experiments demonstrate a good performance of the developed algorithm and its advantage over the one of the best in this field,

the SBMIR algorithm. The choice of the parameters of the AL algorithm is analyzed. The algorithm demonstrates a good convergence rate, and a good accuracy and imaging for the phase and amplitude wave field reconstruction for noiseless and noisy observation data.

6. Acknowledgement

This research is supported by the Academy of Finland, project No. 213462 (Finnish Centre of Excellence program 2006 - 2011), and the post graduate work of Artem Migukin is funded by the Tampere Graduate School in Information Science and Engineering (TISE).

References

1. F. Shen and A. Wang, "Fast-Fourier-transform based numerical integration method for the Rayleigh-Sommerfeld diffraction formula," *Appl. Opt.* **45**, 1102-1110 (2006).
2. J. W. Goodman, *Introduction to Fourier Optics*, 3rd ed. (Roberts & Company, Englewood, 2005).
3. 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).
4. V. Katkovnik, J. Astola, and K. Egiazarian, "Discrete diffraction transform for propagation, reconstruction, and design of wavefield distributions," *Appl. Opt.* **47**, 3481-3493

- (2008).
5. 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).
 6. J. R. Fienup, "Phase retrieval algorithms: A comparison", *Appl. Opt.* **21**, 2758–2769 (1982).
 7. G. Yang, et al. "Gerchberg-Saxton and Yang-Gu algorithms for phase retrieval in a nonunitary transform system: a comparison," *Appl. Opt.* **33**, 209-218 (1994).
 8. D. Russell Luke, J. Burke, and R.Lyon, "Optical Wavefront Reconstruction: Theory and Numerical Methods," *SIAM Review*, Society for Industrial and Applied Mathematics **44**, 169–224 (2002).
 9. J. Burke and D. Russell Luke, "Variational analysis applied to the problem of optical phase retrieval," *SIAM J. on Control and Optim.* **42**, 576–595 (2003).
 10. Heinz H. Bauschke, Patrick L. Combettes, and D. Russell Luke, "Phase retrieval, error reduction algorithm, and Fienup variants: a view from convex optimization," *J. Opt. Soc. Am. A* **19**, 1334-1345 (2002)
 11. A. Migukin, V. Katkovnik, and J. Astola, "3D Wave field phase retrieval from multi-plane observations," *IEEE Conf. 3DTV-CON*, 2010.
 12. Ross W. Deming, "Phase retrieval from intensity-only data by relative entropy minimization," *J. Opt. Soc. Am. A* **24**, 3666-3679 (2007)
 13. F. Soldovieria, R. Deming, R. Pierric, "An improved version of the relative entropy minimization approach for the phase retrieval problem," *Int. J. Electron. Commun. (AEU)* **64**, 56–65.(2010).

14. D. L. Misell, "A method for the solution of the phase problem in electron microscopy," J. Phys. D **6**, L6–L9 (1973).
15. A. Migukin, V. Katkovnik, and J. Astola, "Multiple plane phase retrieval based on inverse regularized imaging and discrete diffraction transform," AIP Conf. Proc. **1236**, ICAPMMOI, 2010.
16. 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).
17. 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).
18. P. Almero, 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).
19. M. Guizar-Sicairos and J. R. Fienup, "Phase retrieval with transverse translation diversity: a nonlinear optimization approach," Opt. Express **16**, 7264-7278 (2008).
20. Xiaojun-Hu , Shengyi-Li , and Yulie-Wu , "Resolution-enhanced subpixel phase retrieval method," Appl. Opt. **47**, 6079-6087 (2008).
21. Gregory R. Brady, Manuel Guizar-Sicairos, and James R. Fienup, "Optical wavefront measurement using phase retrieval with transverse translation diversity," Opt. Express **17**, 624-639 (2009)
22. L. Rudin, S. Osher, and E. Fatemi, "Nonlinear algorithms," Phys. D. **60**, 259-268 (1992).
23. M. Bertero and P. Boccacci. *Introduction to inverse problems in imaging*. (IOP Publish-

ing Ltd, 1998).

24. A.N. Tikhonov and V.Y. Arsenin. *Solution of Ill-Posed Problems* (Wiley, 1977).
25. M. R. Hestenes, "Multiplier and Gradient Methods," JOTA. **4**, 303-320 (1969).
26. M. J. D. Powell, "A method for nonlinear constraints in minimization problems," in *Optimization*, R. Fletcher, ed. (Academic Press, 1969), pp. 283 - 298.
27. J.-F. Cai, S. OSHER, and Z. Shen, "Split Bregman methods and frame based image restoration," SIAM J. Multiscale Model. Simul. **8**, 337-369 (2009).
28. A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," SIAM J. Imaging Sci. **2**, 183-202 (2009).
29. M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," IEEE Trans. Image Proc., 2010, submitted.
30. M. Figueiredo, J. Bioucas-Dias, and R. Nowak, "Majorization-minimization algorithms for wavelet based image restoration," IEEE Trans. Image Proc., **16**, 2980–2991 (2007).
31. J. Dias and J. Leitão, "The ZpM algorithm for interferometric image reconstruction in SAR/SAS", IEEE Trans. Image Proc., **11**, 408-422 (2002).
32. V. Katkovnik and J. Astola, "Wave field reconstruction from multiple intensity observations in 4f optical setup with phase modulation in Fourier focal plane ," Appl. Opt., 2010, submitted.

List of Figure Captions

Fig. 1. Multiple plane wave field reconstruction scenario: $u_0[k]$ and $u_r[k]$ are discrete complex amplitudes at the object and measurement planes respectively, $r = \overline{1, K}$.

Fig. 2 The qualitative comparison of the amplitude reconstruction imaging (*logo* test image) for $\sigma = 0$, *AM*, different number of measurement planes K and distances z_1 . The AL algorithm parameters are $\{\gamma, \alpha, \tilde{\mu}\} = \{10, 1, 0.005\}$.

Fig. 3. The convergence rate of the AL algorithm for the complex-valued wave field reconstruction at the object and sensor planes, obtained using the AL algorithm for $\sigma = 0$, *AM* of the binary test-image *logo*, $d = 2$, $\{\gamma, \alpha, \tilde{\mu}\} = \{10, 1, 0.005\}$. For the sensor planes the mean values and the standard deviations of *RMSE* over $K = 5$ for the amplitude and phase are calculated.

Fig. 4. The convergence rate of the AL algorithm for the complex-valued wave field reconstruction at the object and sensor planes, obtained using the AL algorithm for $\sigma = 0.05$, *PM* of the binary test-image *chessboard*, $\{\gamma, \alpha, \tilde{\mu}\} = \{10, 1, 0.01\}$. For the sensor planes the mean values and the standard deviations of *RMSE* over $K = 5$ for the amplitude and phase are

calculated.

Fig. 5. The object phase reconstruction accuracy via $RMSE$ for $\sigma = 0.05$, $\{\gamma, \alpha, \tilde{\mu}\} = \{10, 1, 0.01\}$ and PM of the binary test image *chessboard* and for different distances Δ_z .

Fig. 6. The AL phase reconstruction $\sigma = 0$, PM , $\{\gamma, \alpha, \tilde{\mu}\} = \{10, 1, 0.001\}$: the first row (a)-(c) $K = 3$, $\Delta_z = 0.4 \cdot z_f$; the second row (d)-(f) $K = 5$, $\Delta_z = 0.3 \cdot z_f$; the third row (g)-(i) $K = 10$, $\Delta_z = 0.15 \cdot z_f$. The images in columns from left to the right corresponds to different distances: $d = \{0.5, 1, 2\}$ respectively, $z_1 = d \cdot z_f$, $z_f = 10.8mm$.

Fig. 7. The comparison of the object wave field reconstruction for $\sigma = 0$, $K = 5$, AM of the test-image *lena*. The first column is obtained by AL, $\{\gamma, \alpha, \tilde{\mu}\} = \{10, 1, 5 \cdot 10^{-4}\}$: (a) amplitude reconstruction, $RMSE(|u_0|) = 0.041$, (c) phase reconstruction, $RMSE(\phi_0) = 0.091$. The second column is obtained by SBMIR: (b) amplitude reconstruction, $RMSE(|u_0|) = 0.08$, (d) phase reconstruction, $RMSE(\phi_0) = 0.28$.

Fig. 8. The convergence rate of the AL and SBMIR algorithm for the amplitude reconstruction (*lena*), the case of AM .

Fig. 9. The comparison of the object wave field reconstruction for $\sigma = 0.05$, $K = 5$, PM of the test-image *chessboard*. The first column is obtained by AL, $\tilde{\mu} = 0.01$: (a) amplitude reconstruction, $RMSE(|u_0|) = 0.23$, (c) phase reconstruction, $RMSE(\phi_0) = 0.26$. The second column is obtained by SBMIR: (b) amplitude reconstruction, $RMSE(|u_0|) = 0.35$, (d) phase reconstruction, $RMSE(\phi_0) = 0.58$.

Fig. 10. The convergence rate of the AL and SBMIR algorithm for the phase reconstruction (*chessboard*), the case of PM .

Fig. 11. The cross-sections for the reconstruction of the smooth object phases, PM , $K = 5$, $\sigma = 0.05$, 1000 iterations: (dotted curve) the normalized true image (*Mexican Hat*, see Eq.(28)) and the normalized unwrapped reconstructions of the phase, obtained by (dashed curve) the AL algorithm, $\tilde{\mu} = 0.01$, $RMSE(\phi_0) = 0.187$, (solid curve) the SBMIR algorithm, $RMSE(\phi_0) = 0.511$.

Fig. 12. The convergence rate of the AL and SBMIR algorithm for the phase reconstruction (*Mexican Hat*).

List of Table Captions

Table 1. The quantitative comparison of the amplitude reconstruction, $RMSE(|u_0|)$ for the test, presented in Fig. 2.

Table 2. The quantitative comparison of the phase reconstruction (via $RMSE(\phi_0)$) for the test, presented in Fig. 6.

Table 3. The computational time (in sec) for 100 iterations of the AL algorithm.

Table 4. The computational time (in sec) for calculation of K wave field propagation operators with double size $F-DDT$.

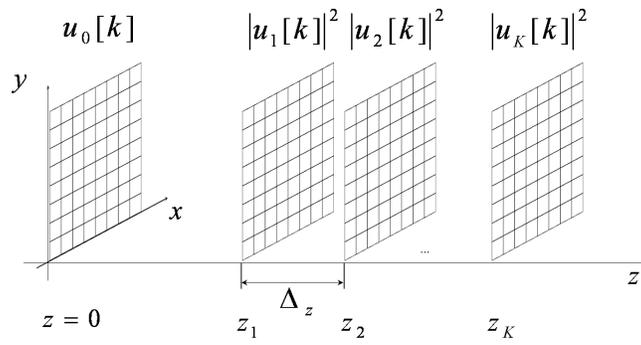


Fig. 1. Multiple plane wave field reconstruction scenario: $u_0[k]$ and $u_r[k]$ are discrete complex amplitudes at the object and measurement planes respectively, $r = \overline{1, K}$.

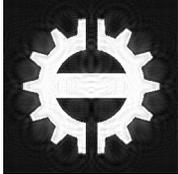
$d \backslash K$	3	5	10
0.5			
1			
1.5			
2			
3			

Fig. 2. The qualitative comparison of the amplitude reconstruction imaging (*logo* test image) for $\sigma = 0$, AM , different number of measurement planes K and distances z_1 . The AL algorithm parameters are $\{\gamma, \alpha, \tilde{\mu}\} = \{10, 1, 0.005\}$.

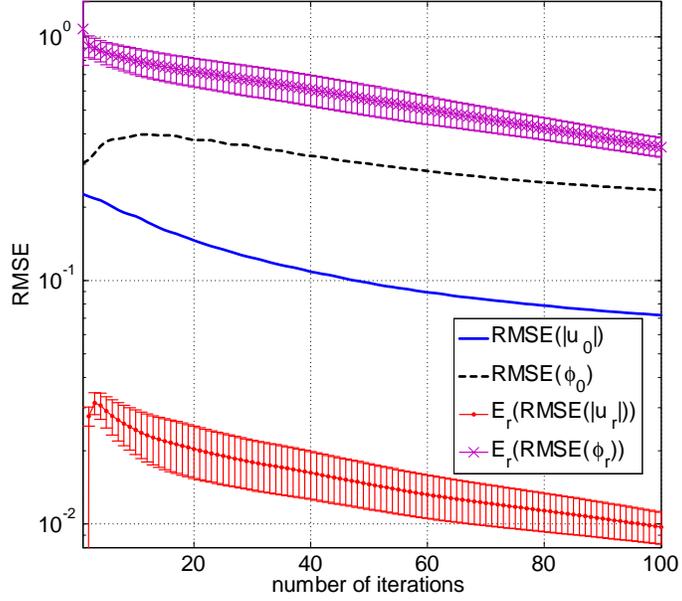


Fig. 3. The convergence rate of the AL algorithm for the complex-valued wave field reconstruction at the object and sensor planes, obtained using the AL algorithm for $\sigma = 0$, AM of the binary test-image *logo*, $d = 2$, $\{\gamma, \alpha, \tilde{\mu}\} = \{10, 1, 0.005\}$. For the sensor planes the mean values and the standard deviations of $RMSE$ over $K = 5$ for the amplitude and phase are calculated.

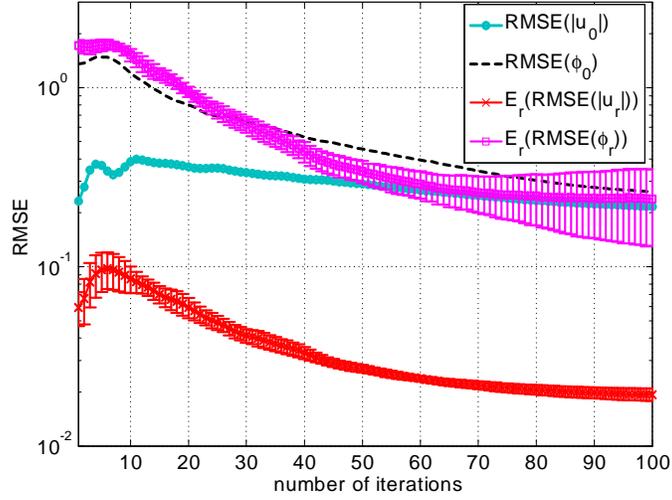


Fig. 4. The convergence rate of the AL algorithm for the complex-valued wave field reconstruction at the object and sensor planes, obtained using the AL algorithm for $\sigma = 0.05$, PM of the binary test-image *chessboard*, $\{\gamma, \alpha, \tilde{\mu}\} = \{10, 1, 0.01\}$. For the sensor planes the mean values and the standard deviations of $RMSE$ over $K = 5$ for the amplitude and phase are calculated.

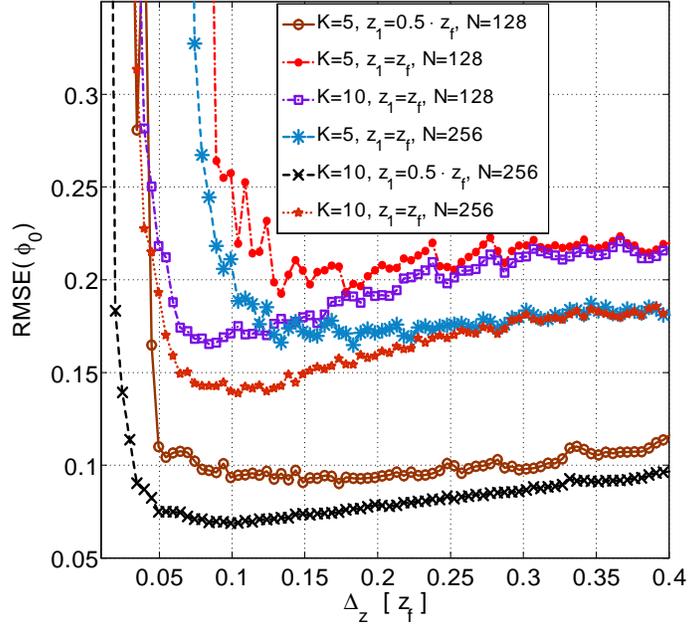


Fig. 5. The object phase reconstruction accuracy via $RMSE$ for $\sigma = 0.05$, $\{\gamma, \alpha, \tilde{\mu}\} = \{10, 1, 0.01\}$ and PM of the binary test image *chessboard* and for different distances Δ_z .

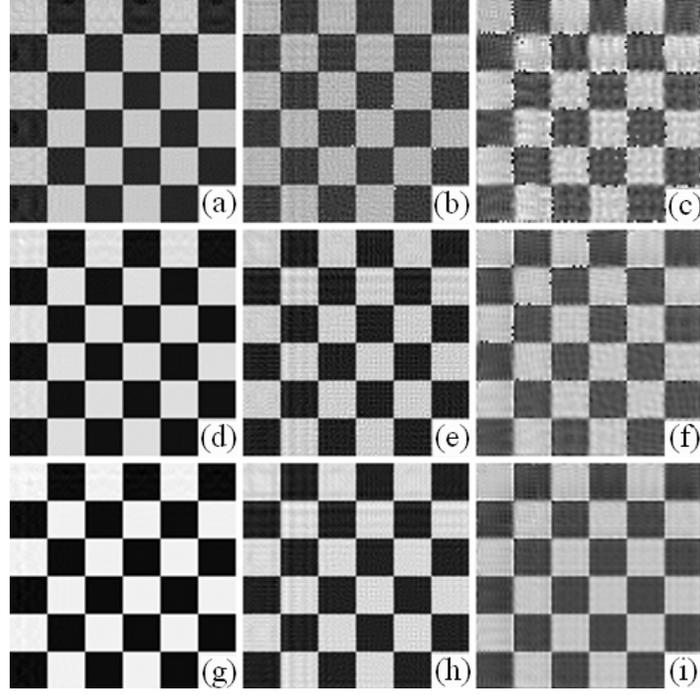


Fig. 6. The AL phase reconstruction $\sigma = 0$, PM , $\{\gamma, \alpha, \tilde{\mu}\} = \{10, 1, 0.001\}$: the first row (a)-(c) $K = 3$, $\Delta_z = 0.4 \cdot z_f$; the second row (d)-(f) $K = 5$, $\Delta_z = 0.3 \cdot z_f$; the third row (g)-(i) $K = 10$, $\Delta_z = 0.15 \cdot z_f$. The images in columns from left to the right corresponds to different distances: $d = \{0.5, 1, 2\}$ respectively, $z_1 = d \cdot z_f$, $z_f = 10.8mm$.

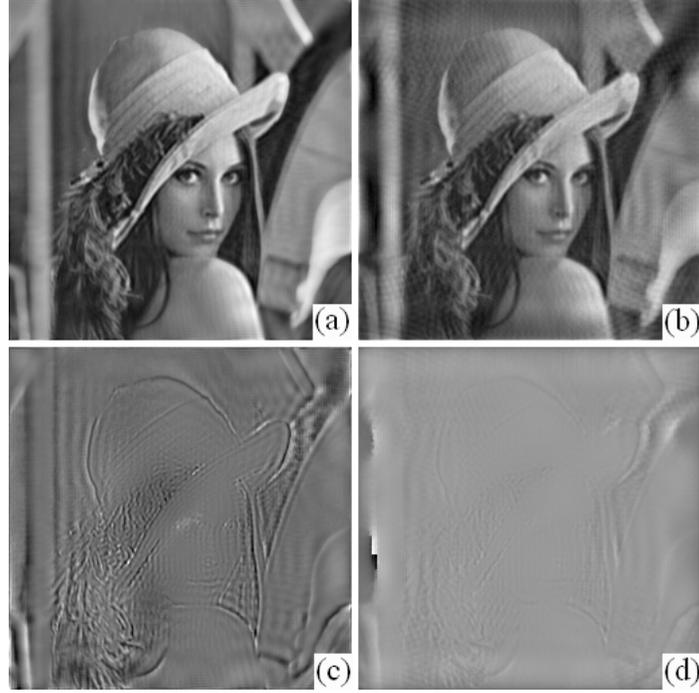


Fig. 7. The comparison of the object wave field reconstruction for $\sigma = 0$, $K = 5$, AM of the test-image *lena*. The first column is obtained by AL, $\{\gamma, \alpha, \tilde{\mu}\} = \{10, 1, 5 \cdot 10^{-4}\}$: (a) amplitude reconstruction, $RMSE(|u_0|) = 0.041$, (c) phase reconstruction, $RMSE(\phi_0) = 0.091$. The second column is obtained by SBMIR: (b) amplitude reconstruction, $RMSE(|u_0|) = 0.08$, (d) phase reconstruction, $RMSE(\phi_0) = 0.28$.

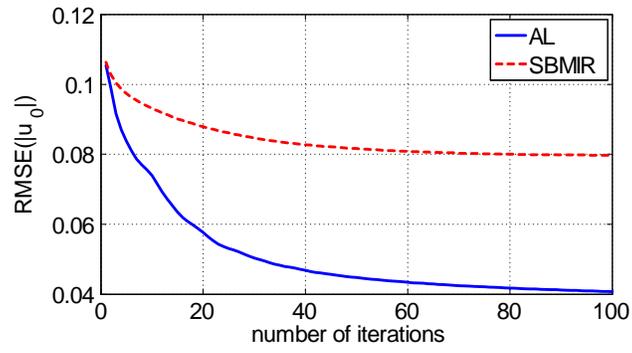


Fig. 8. The convergence rate of the AL and SBMR algorithm for the amplitude reconstruction (*lena*), the case of *AM*.

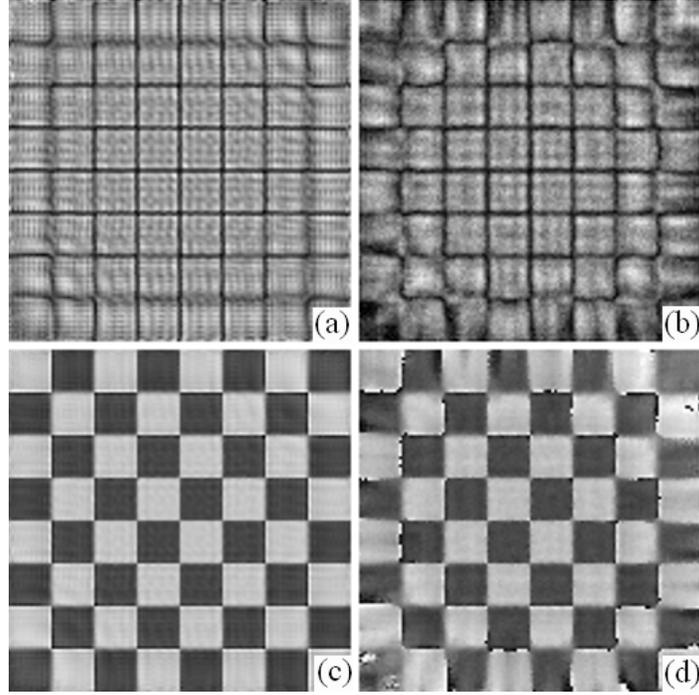


Fig. 9. The comparison of the object wave field reconstruction for $\sigma = 0.05$, $K = 5$, PM of the test-image *chessboard*. The first column is obtained by AL, $\tilde{\mu} = 0.01$: (a) amplitude reconstruction, $RMSE(|u_0|) = 0.23$, (c) phase reconstruction, $RMSE(\phi_0) = 0.26$. The second column is obtained by SBMIR: (b) amplitude reconstruction, $RMSE(|u_0|) = 0.35$, (d) phase reconstruction, $RMSE(\phi_0) = 0.58$.

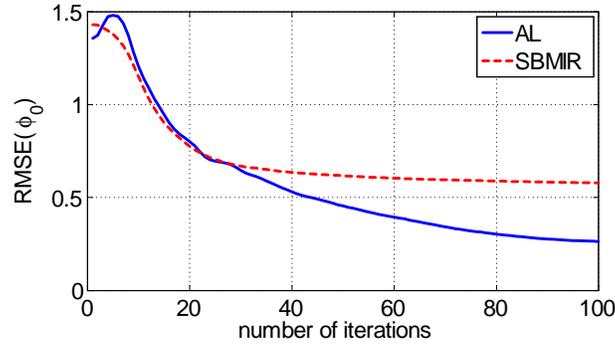


Fig. 10. The convergence rate of the AL and SBMIR algorithm for the phase reconstruction (*chessboard*), the case of PM .

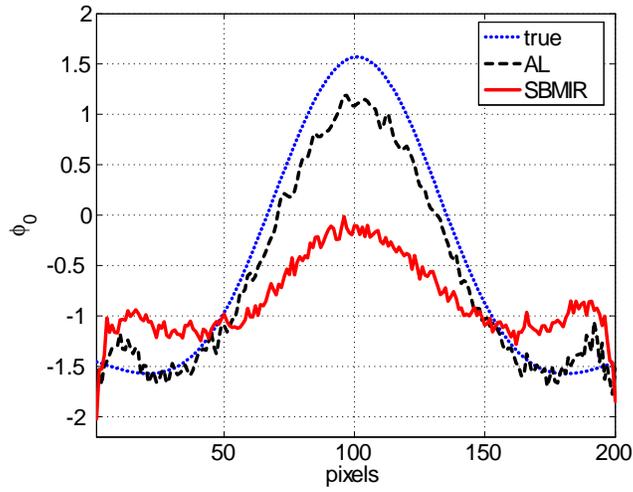


Fig. 11. The cross-sections for the reconstruction of the smooth object phases, PM , $K = 5$, $\sigma = 0.05$, 1000 iterations: (dotted curve) the normalized true image (*Mexican Hat*, see Eq.(28)) and the normalized unwrapped reconstructions of the phase, obtained by (dashed curve) the AL algorithm, $\tilde{\mu} = 0.01$, $RMSE(\phi_0) = 0.187$, (solid curve) the SBMIR algorithm, $RMSE(\phi_0) = 0.511$.

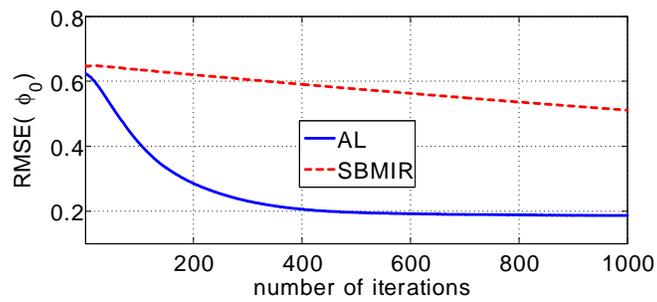


Fig. 12. The convergence rate of the AL and SBMIR algorithm for the phase reconstruction (*Mexican Hat*).