PHASE RETRIEVAL FROM MULTIPLE PLANE OBSERVATIONS: CONSTRAINED VARIATIONAL FORMULATION AND AUGMENTED LAGRANGIAN RECURSIVE ALGORITHM

Artem Migukin, Vladimir Katkovnik, and Jaakko Astola

Department of Signal Processing, Tampere University of Technology (TUT),
Korkeakoulunkatu 10, P.O.Box 553, FI-33720 Tampere, Finland
artem.migukin@tut.fi, vladimir.katkovnik@tut.fi and jaakko.astola@tut.fi

ABSTRACT

A new recursive augmented Lagrangian (AL) algorithm is presented for the module and phase reconstruction of a 3D wave field from intensity-only measurements on two or more sensor planes at different axial positions. The wave field reconstruction is framed as a constrained nonlinear optimization problem allowing to involve prior information on a wave field of interest. The main goal is to design the algorithm which is more accurate than the conventional ones such as the well-known Gerchberg–Saxton algorithms and their multiple modifications [1]-[5]. A numerical study of the proposed AL algorithm for various types of the object modulation shows a better accuracy of wave field reconstruction and better imaging for different setup parameters comparing with the successive iterative method, originated in [3].

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 between the parallel object and the $r-th$ observation planes, and K is a number of the observation planes. In discrete modeling all continuous variables x are replaced by the digital ones with the following replacements of the continuous distributions by their discrete counterparts: $u_0(x) \rightarrow u_0[k]$, $u_r(x) \rightarrow u_r[k]$ with integer arguments k.

The discrete intensity observations are given in the form

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

where the wave field intensity (power) is measured with an additive random errors $\varepsilon_r[k]$. We assume that the noise is zero-mean Gaussian, $\varepsilon_r[k] \sim \mathcal{N}(0, \sigma_r^2)$.

The problem is to reconstruct pixelated complex-valued wave field distributions $u_o[k]$ and $u_r[k]$ at the object and sensor planes from the data (1). The problem setup is illustrated in Fig.1. Δ_z is the distance between the observation planes.

This distribution (image) restoration is known as an inverse problem. For simplicity, we denote images as vectors in R^n by concatenating their column and use bold lower case characters for these vectors. Then, the 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 $r-th$ image (sensor) plane as

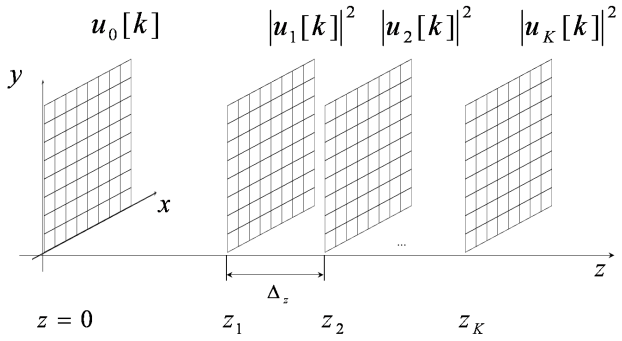

Figure 1. Multiple plane wave field reconstruction scenario: $u_o[k]$ and $u_r[k]$ are discrete complex amplitudes at the object and measurement planes respectively, $r = \overline{1,K}$

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

where $\mathbf{A}_r$ is the propagation operator (complex-valued matrix) 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 equations. The operator $\mathbf{A}_r$ in (2) is specified by discretization of this modeling. It can be convolutional single or double size model, angular spectrum decomposition (ASD) [6], or recent discrete diffraction transforms in matrix (M-DDT) [7] or frequency (F-DDT) forms [8]. 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 $\mathbf{u}_r$.

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

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

where the modulus $|\cdot|$ and square $|\cdot|^2$ are the point-wise operations applied to the elements of the corresponding vectors.

2. VARIATIONAL WAVE FIELD RECONSTRUCTION

In this article, we do not follow any variant of Gerchberg-Saxton or Fienup's error-reduction algorithms [1], [2], [9] but rather apply the maximum likelihood style approach. For the Gaussian noise distribution and the observation model (3) it results in the following criterion

$$J = \sum_r \frac{1}{2\sigma_r^2} ||\mathbf{o}_r - |\mathbf{u}_r|^2||_2^2 + \mu \cdot pen(\mathbf{u}_0), \quad (4)$$

where the norm $||\cdot||_2^2$ is Euclidean and the power in $|\mathbf{u}_r|^2$ is an element-wise operation. The first summand in (4) is obtained as the main term of the minus logarithm of the Gaussian likelihood function corresponding to the observation model (3), and the second summand is the penalty (regularization) including the prior on the object distribution $\mathbf{u}_0$ to be reconstructed.

The wave field reconstruction is formulated as the following optimization problem:

$$\hat{\mathbf{u}}_0 = \arg\min_{\mathbf{u}_0} \sum_r \frac{1}{2\sigma_r^2} ||\mathbf{o}_r - |\mathbf{u}_r|^2||_2^2 + \quad (5)$$
$$+\mu \cdot pen(\mathbf{u}_0)$$
$$\text{subject to } \mathbf{u}_r = \mathbf{A}_r \mathbf{u}_0.$$

with the only unknown variable $\mathbf{u}_0$ and the distributions $\mathbf{u}_r$ calculated according to the forward propagation models (2).

The parameter μ in (5) defines a balance between the accuracy of the observation fitting and the prior. 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 over-smoothed 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.

3. AUGMENTED LAGRANGIAN ALGORITHM

The augmented Lagrangian (AL) method, introduced independently by Hestenes [10] and Powell [11] is now classical for the minimization of functionals in presence of linear equality constraints.

The augmented Lagrangian corresponding to (5) is of the form

$$L(\mathbf{u}_0, \{\mathbf{u}_r\}, \{\mathbf{\Lambda}_r\}) = \quad (6)$$
$$\sum_r \frac{1}{\sigma_r^2} [\frac{1}{2} ||\mathbf{o}_r - |\mathbf{u}_r|^2||^2 + \frac{1}{\gamma_r} ||\mathbf{u}_r - \mathbf{A}_r \cdot \mathbf{u}_0||^2 +$$
$$+ \frac{2}{\gamma_r} \mathrm{Re}\{\mathbf{\Lambda}_r^{*T} (\mathbf{u}_r - \mathbf{A}_r \cdot \mathbf{u}_0)\}] + \mu ||\mathbf{u}_0||_2^2.$$

The Lagrangian based optimization is associated with the saddle problem, which requires minimization on $\mathbf{u}_0$, $\{\mathbf{u}_r\}$ and maximization on the vector of the Lagrange multipliers $\{\mathbf{\Lambda}_r\}$. The parameters γ_r are positive.

In AL both the linear and quadratic terms in (6) correspond to the linear constrains $\mathbf{u}_r - \mathbf{A}_r \cdot \mathbf{u}_0 = 0$. If we keep only the quadratic terms the augmented Lagrangian becomes the penalty criterion, which assumes that the penalty coefficients $\frac{1}{\gamma_r}$ are large. As a rule it leads to computational difficulties because this criterion can be very ill-conditioned. If we keep only the linear terms the augmented Lagrangian becomes the standard Lagrangian. However, the saddle-point of this standard Lagrangian is unstable. It results in the problems with numerical solutions. The stability of the saddle-point of the augmented Lagrangian keeping both the linear and quadratic terms is one of the principal advantages of this criterion.

The proposed algorithm corresponding to a recursive optimization of the criterion $L(\mathbf{u}_0, \{\mathbf{u}_r\}, \{\mathbf{\Lambda}_r\})$ is combined from the following successive steps :

AL Phase Algorithm (7)

1. Set $t = 0$ (initialization), $\mathbf{u}_{0,0}$, $\mathbf{\Lambda}_{r,0}$,
2. Repeat, $t = 0, 1, ...$ for $r = \overline{1, K}$
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. $\mathbf{u}_{0,t+1} = (\sum_r \frac{1}{\gamma_r \sigma_r^2} \mathbf{A}_r^{*T} \mathbf{A}_r + \mu \cdot \mathbf{I}_{n \times n})^{-1} \times$
$\sum_r \frac{1}{\gamma_r \sigma_r^2} \mathbf{A}_r^{*T} (\mathbf{u}_{r,t+1} + \mathbf{\Lambda}_{r,t})$,
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

In other words the algorithm (7) calculate the estimate $\hat{\mathbf{u}}_0 = \max_{\{\mathbf{\Lambda}_r\}} \min_{\mathbf{u}_0, \{\mathbf{u}_r\}} L(\mathbf{u}_0, \{\mathbf{u}_r\}, \{\mathbf{\Lambda}_r\})$. In Step 4 $\mathcal{G}(\mathbf{o}_r(k), \mathbf{u}_{r,t}(k), \mathbf{\Lambda}_{r,t}(k))$ is a solution of the minimization $\min_{\{\mathbf{u}_r\}} L(\mathbf{u}_0, \{\mathbf{u}_{r,t}\}, \{\mathbf{\Lambda}_{r,t}\})$. Step 5 updates the reconstruction of the object distribution $\mathbf{u}_{0,t+1}$, used in Step 3 for prediction of the wave field distribution at the $r - th$ plane. Step 6 defines the updates of the Lagrange multipliers.

4. SIMULATION EXPERIMENTS

The numerical experiments are performed for the amplitude ($AM, \phi_0[k] = 0$) and phase modulation ($PM, |u_0[k]| = 1$) of an object distribution of the form $u_0[k] = |u_0[k]| \cdot \exp(j \cdot \phi_0[k])$. The results are presented for $t = 100$ iterations of the algorithm.

In our simulation study we consider the selection of the algorithm parameters $\{\gamma_r, \alpha_r, \mu\}$ and analyze the performance of the AL algorithm. As $N \times M$ test-images we use the gray-scale *lena*, binary *chessboard* and the smooth *Mexican Hat* of the corresponding sizes $256 \times 256, 128 \times 128, 200 \times 200$, respectively. Then $n = N \cdot M$. The last test image is a discrete sample from

$$u_0(x) = -\frac{4}{3\sqrt{\pi}} \cdot (||x||^2 - 2) \cdot \exp(-\frac{1}{2} ||x||^2), \quad (8)$$

for the 2D discrete grid $-6 \leq (x_1, x_2) \leq 6$ with the discretization step equal to 0.025.

We assume that sensor and object distributions are pixelated with square pixels $\Delta \times \Delta$ and 100% fill factor, $\Delta = 6.7\mu m$. In the numerical experiments the following setup parameters are used: the wavelength $\lambda = 532nm$ (corresponding to a Nd:YAG green laser), the distance between the observation planes $\Delta_z = 1\ mm$ (as in [4] with the best results of the successive iterative method). A significant role in the wave field reconstruction accuracy plays the number of measurement planes K, which varies in the interval $[2, 15]$. The "in-focus" distance is calculated as $z_f = N \cdot \Delta^2/\lambda$ (for square $N \times N$ test images), and the distance between the object and the first measurement planes $z_1 = d \cdot z_f$ is expressed through the "in-focus" distance z_f with the scale parameter d. In the presented results $d = 2$. Effects of the measurement quantization in the observation planes are out of the scope in this work, and we assume that double precision data are given. The wave field reconstruction accuracy is calculated via the root-mean-square error ($RMSE$).

It is found that $\mathbf{u}_{0,0} = \frac{1}{2} \cdot \mathbf{1}_{n\times 1}$ is a good initial guess for both the amplitude $u_0[k] = |u_0[k]| \cdot \exp(j \cdot 0)$ and the phase modulation $u_0[k] = 1 \cdot \exp(j \cdot \phi_0[k])$ of the object distribution. We consider the selection of an appropriate algorithm parameters $\{\gamma,\alpha,\mu\}$, assuming that $\alpha_r = \alpha$, $\gamma_r = \gamma$, $\sigma_r^2 = \sigma^2$ for all r. Then, for noisy data $\mu \to \tilde{\mu} = \mu \cdot \sigma^2$, for noiseless data $\tilde{\mu} = \mu$, and Step 5 in (7) takes the form:

$$\mathbf{u}_{0,t+1} = [\sum_r \mathbf{A}_r^{*T}\mathbf{A}_r + \tilde{\mu}\gamma \cdot \mathbf{I}_{n\times n}]^{-1} \times \sum_r \mathbf{A}_r^{*T}(\mathbf{u}_{r,t+1} + \mathbf{\Lambda}_{r,t}) \quad (9)$$

4.1. Selection of the algorithm parameters

In this section we present the recommended values of the algorithm parameters $\{\gamma,\alpha,\mu\}$ which are found being acceptable for both AM and PM object distributions, for different distances and for noisy and noiseless data.

The step-size $\alpha > 0$ should be close, but not larger than 1. Its optimal values vary a little depending on other experiment parameters. The recommended value is $\alpha = 1$.

The regularization parameter $\tilde{\mu}$ should be increased for larger values of d. It is found also that for PM the regularization parameter $\tilde{\mu}$ should be in order larger than for AM. In general, it is convenient to use $\tilde{\mu}$ starting from values $10^{-3}/10^{-2}$ in case of noisy data (or $\tilde{\mu} = 5 \cdot 10^{-5}/5 \cdot 10^{-3}$ in case of noiseless data) for AM/PM, respectively.

It is recommended to use $\gamma = 10 \div 20$ and take larger values for $z_1 > z_f$. In our experiments we fix $\gamma = 10$ for all types of the object distributions.

4.2. Convergence and wave field reconstruction accuracy

Let us consider the algorithm performance both for noiseless ($\sigma = 0$) or noisy data ($\sigma = 0.05$) for the parameters $\{\gamma,\alpha,\mu\}$ selected as discussed above. The observations $\mathbf{o}_r$ are always generated using Discrete Diffraction Transform (DDT) for the forward propagation modeling because DDT modeling is precise for pixelated sensor and object distributions.

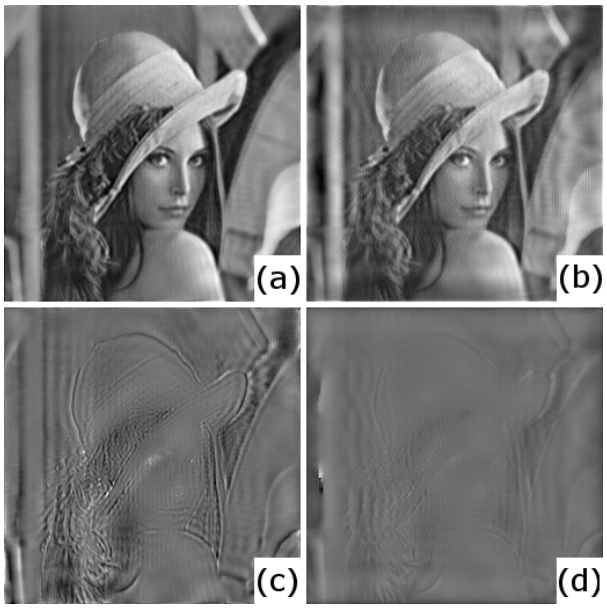

Figure 2. The comparison of the object wave field reconstruction for $\sigma = 0$, $K = 5$, AM. The first column is obtained by AL, $\tilde{\mu} = 10^{-4}$: (a) amplitude reconstruction $|\hat{u}_0|$, $RMSE = 0.046$, (c) phase reconstruction $\hat{\phi}_0$, $RMSE = 0.1$. The second column is obtained by SBMIR: (b) amplitude reconstruction $|\hat{u}_0|$, $RMSE = 0.088$, (d) phase reconstruction $\hat{\phi}_0$, $RMSE = 0.24$.

The comparison is produced versus the single-beam multiple-intensity phase reconstruction (SBMIR) algorithm originated in [3]. The unknown object wave field is reconstructed as a complex-valued distribution assuming that we do not have prior information on the object type, so the amplitude and phase of the modulated object are recovered.

4.2.1. Gray-scale "lena" test image

Imaging of the reconstructed amplitude $|\hat{u}_0|$ and phase $\hat{\phi}_0$ at the object plane is shown in Fig.2. These results are given for the amplitude modulation of the object distribution. The accuracy of the AL algorithm for both amplitude and phase reconstructions is approximately twice better in $RMSE$ values than the accuracy of 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 damaging visualization. There are no such artifacts in the AL imaging.

4.2.2. Binary "chessboard" test chart

If the number of measurement planes K is small (say, $K = 3$ or 5) an advantage of the AL algorithm is clear.

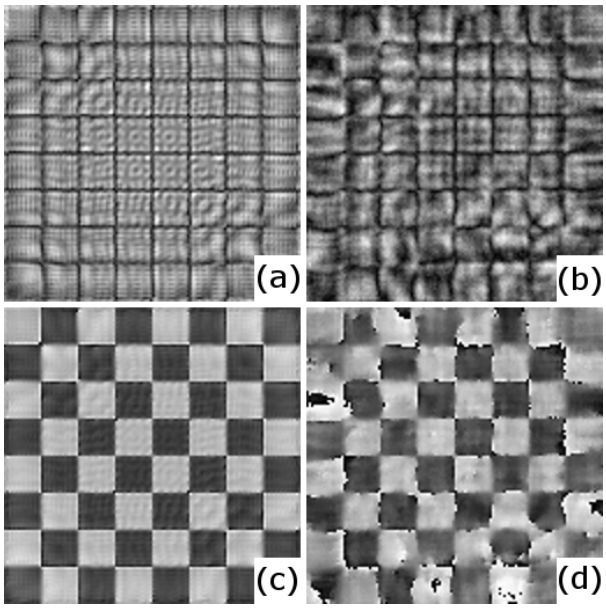

Figure 3. The comparison of the object wave field reconstruction for $\sigma = 0.05$, $K = 5$, PM. The first column is obtained by AL, $\tilde{\mu} = 0.01$: (a) amplitude reconstruction $|\hat{u}_0|$, $RMSE = 0.26$, (c) phase reconstruction $\hat{\phi}_0$, $RMSE = 0.36$. The second column is obtained by SBMIR: (b) amplitude reconstruction $|\hat{u}_0|$, $RMSE = 0.34$, (d) phase reconstruction $\hat{\phi}_0$, $RMSE = 0.9$.

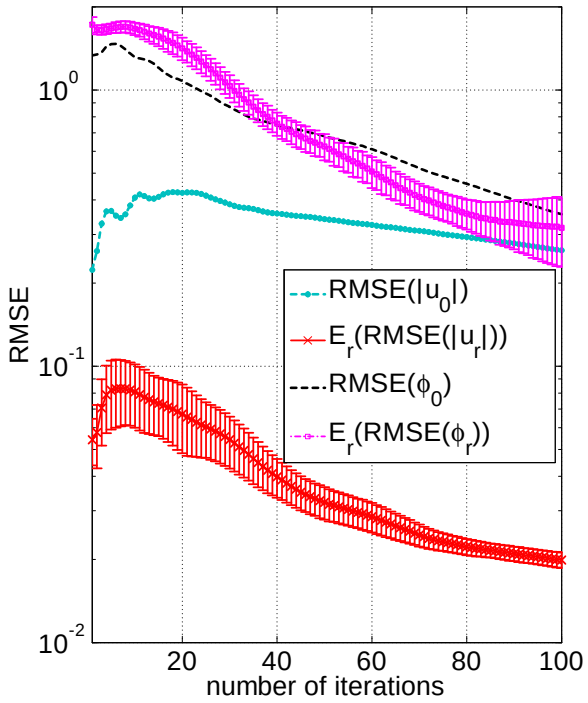

Figure 4. The convergence rate of the AL algorithm for the complex-valued wave field reconstruction at the object and sensor planes, obtained using the AL algorithm for $\sigma = 0.05$, PM, $\tilde{\mu} = 0.01$. For the sensor planes the mean values and the standard deviations of $RMSE$ over $K = 5$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

For example, in Fig.3 we compare the wave field reconstruction for the phase modulated binary test image. We consider the case of noisy data ($\sigma = 0.05$) for $K = 5$. It is seen that the AL algorithm demonstrates quite clear and sharp results (left column), while the SBMIR reconstruction is blurred and significantly destroyed (right column). The $RMSE$ values for the amplitude $|\hat{u}_0|$ and phase $\hat{\phi}_0$ are 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.

The convergence rates of the AL algorithm for the object and sensor planes are illustrated in Fig.4. The accuracy of reconstruction is considered with calculation of $RMSE$ for the phases and amplitudes separately. For the measurement planes the mean value and the standard deviation of $RMSE$ are calculated over the K planes. The solid curves, which are 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. The vertical intervals around these curves show the standard deviations of the corresponding $RMSEs$. The convergence in both the mean value and the standard deviation for $RMSE(|u_r|)$ confirms the convergence of the amplitude reconstruction at the sensor planes. For the phase the situation is more complex, because there is convergence in the mean value of $E_r(RMSE(\phi_r))$ while we can see the divergence in the corresponding standard deviation.

It is found that 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 the accuracy in $RMSE$ values by approximately $70\%/30\%$ for the reconstruction of the amplitude/phase, respectively. Five more planes (from $K = 5$ to $K = 10$) results in the improvement by approximately $40\%/25\%$ for AM/PM respectively. A further increasing of K does not give essential effects.

4.2.3. "Mexican Hat"

Note that both "*lena*" and "*chessboard*" are examples of discontinuous non-smooth distributions. In contrast to it here we consider experiments produced for the continuous smooth test-image "*Mexican Hat*", which is used to define the phase for PM of the object distribution. In Fig. 5 we show the normalized true phase and two reconstructions, obtained by the AL and SBMIR algorithms, $\Delta_z = 3\,mm$. The advantage of the AL algorithm is obvious. The convergence rate is illustrated in Fig. 6 with the number of iterations up to 1000.

It is shown that the AL algorithm converges much faster than SBMIR and gives essentially better result. The almost full bell and the hollow of *Mexican Hat* are reconstructed by AL (Fig.5(b)). However, there are significant errors on the borders in the AL and SBMIR reconstruc-

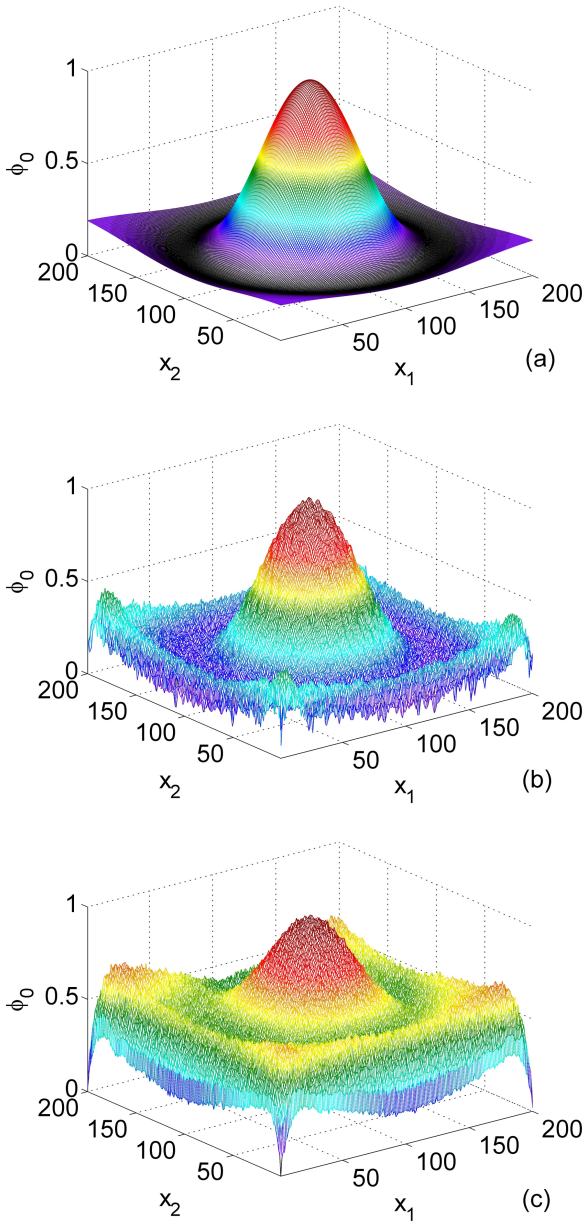

Figure 5. The reconstruction of the smooth object phase PM, $K = 5$, $\sigma = 0.05$, 1000 iterations, $\Delta_z = 3\,mm$: (a) the normalized true image (*Mexican Hat*, see Eq.(8)) and the normalized reconstructions of the phase, obtained by (b) the AL algorithm, $\tilde{\mu} = 0.01$, $RMSE = 0.16$, (c) the SBMIR algorithm, $RMSE = 0.43$.

tions. Overall the SBMIR reconstruction is much worse in terms of the accuracy and the convergence rate.

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 and the prior information on the wave field to be recovered.

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. This approach can be further developed for reconstruction of the only amplitude or phase modulated objects.

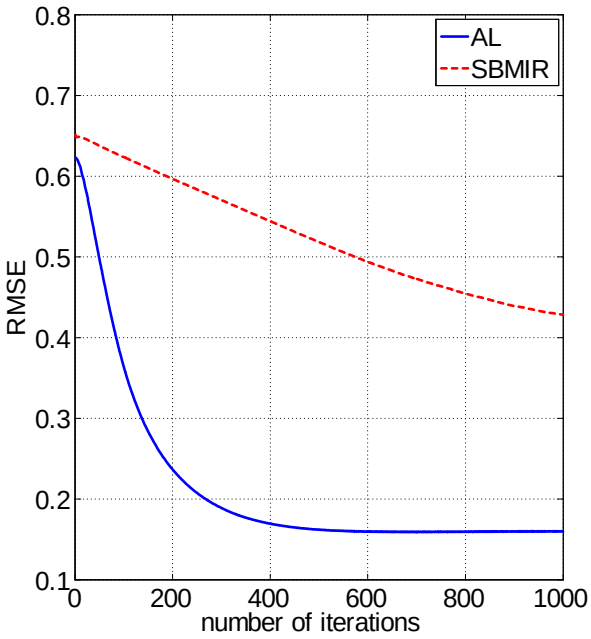

Figure 6. The convergence rate of the AL ($\tilde{\mu} = 0.01$) and SBMIR algorithm for the phase reconstruction (*Mexican Hat*).

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

6. ACKNOWLEDGMENTS

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).

7. REFERENCES

[1] R. W. Gerchberg and W. O. Saxton, "A practical algorithm for the determination of phase from image and diffraction plane pictures," *Optik*, vol. 35, pp. 227-246, 1972.

[2] J. R. Fienup, "Phase retrieval algorithms: a comparison," *Appl. Opt.*, vol.2l, pp. 2758-2769, 1982.

[3] G. Pedrini, W. Osten, and Y. Zhang, “Wave-front reconstruction from a sequence of interferograms recorded at different planes,” *Opt. Lett.* vol.30, pp. 833-835, 2005.

[4] P. Almoro, G. Pedrini and W. Osten, "Complete wavefront reconstruction using sequential intensity measurements of a volume speckle field," *Appl. Opt.*, Vol. 45, pp.8596-8605, 2006.

[5] P. Almoro, A. M. S. Maallo, and S. Hanson, "Fast-convergent algorithm for speckle-based phase retrieval and a design for dynamic wavefront sensing," *Appl. Opt.* 48, pp.1485-1493, 2009

[6] J. W. Goodman, Introduction to Fourier Optics, New York: McGraw-Hill Inc, Second Edition, 1996.

[7] V. Katkovnik, J. Astola, and K. Egiazarian, “Discrete diffraction transform for propagation, reconstruction, and design of wavefield distributions,“ *Appl. Opt.* vol.47, pp. 3481-3493, 2008.

[8] 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.* vol.48, pp. 3407-3423, 2009.

[9] M. Guizar-Sicairos and J. R. Fienup, “Phase retrieval with transverse translation diversity: a nonlinear optimization approach,“ *Opt. Express*, vol.16, pp. 7264-7278, 2008.

[10] M. R. Hestenes, "Multiplier and Gradient Methods," *Computing Method in Optimization Problems 2*, L. A. Zadeh, L. W. Neustadt and A. V. Balakrishnan, eds., Academic Press, New York, 1969.

[11] M. J. D. Powell, "A Method for the Non-Linear Constraints in Minimization Problems," Optimization, R. Fletcher, ed., Academic Press, New York, 1972.