Advanced phase retrieval: augmented Lagrangian based algorithm with sparse regularization of object phase and amplitude

Artem Migukin
Department of Signal Processing, Tampere University of Technology
P.O. Box 527, 33101 Tampere, Finland
artem.migukin@tut.fi

April 27, 2011

Abstract

In this paper, we present a novel iterative phase-retrieval algorithm using a sparse representation of the object amplitude and phase. Sparse modeling is one of the efficient techniques for imaging that allows recovering lost information on the object distribution. The algorithm is derived in terms of constrained maximum likelihood assuming that the wave field reconstruction is performed from a number of noisy intensity-only observations with a zero-mean additive Gaussian noise. The developed algorithm enables the optimal solution for the object distribution reconstruction, and sparse regularization results in advanced reconstruction accuracy. Numerical simulations demonstrate significant enhancement of imaging.

1 Introduction

The conventional sensors detect only the intensity of the light, but the phase is systematically lost in measurements. Phase retrieval is a problem of the phase recovering using a number of intensity observations and some prior on the object. The phase carries important information about the object shape what is necessary for a 3D object imaging and exploited in many areas such as microscopy, astronomy, etc. Moreover, phase-retrieval techniques are often simpler, cheaper and more robust comparing with interferometric ones.

In 1982 Fienup systematize earlier works and introduced some, for now classical, iterative phase-retrieval algorithms [1]: error-reduction, gradient search and input-output methods. Many phase-retrieval methods are developed based on this pioneer work: the estimated magnitudes at the measurement planes are iteratively replaced by ones obtained from the intensity observations. These algorithms differ by the representation of an object wave field.

Contrary to that, Misell's variation of the phase-retieval algorithm [2] operates with data at the measurement planes only, and no connection with the object is used. Recently such successive phase recovering from data at non-focal planes only are shown to be very effective. One of the most efficient successive algorithms is the iterative method known as single-beam multiple-intensity phase reconstruction ($SBMIR$) [3].

We are looking for an optimal wave field reconstruction from a number of intensity observations, and the object estimation is formulated in terms of variational constrained maximum likelihood (ML) approach. The spatial image resolution of the conventional phase-retrieval techniques is limited due to diffraction, what can be the main source of artifacts and image degradation. In order to enhance the imaging quality and recover lost information, in this work we use the newly-developed compressive sensing technique for the variational image reconstruction. The object is assumed to be sparse, and its amplitude and phase are separately decomposed using very specific basis functions called as BM3D-frames [4]. The proposed phase-retrieval algorithm is derived as a solution of the ML optimization problem using the BM3D-frame based sparse approximation of the object distribution.

2 Wave field propagation model

We consider a multi-plane wave field reconstruction scenario: a planar laser beam illuminates an object, and the result of the wave field propagation is detected on a sensor at different distances from the object z_l, at various measurement (sensor) planes parallel to the object plane. Here $z_l = z_1 + (l-1) \cdot \Delta_z$, $l = 1, ...L$, where z_1 is the distance from the object to the first measurement plane, Δ_z is the distance between the measurement planes, and L is a number these planes. We assume that the wave field distributions at the object and sensor planes are pixel-wise invariant. In such a discrete-to-discrete model, the forward wave field propagation from the object to the $l-$th sensor plane can be presented as follows:

$$\mathbf{u}_{z_l} = A_l \cdot \mathbf{u}_0,\ l = 1, ...L, \tag{1}$$

where $\mathbf{u}_0$ and $\mathbf{u}_{z_l}$ are $\mathbb{C}^{n\times 1}$ vectors, constructed by columns concatenating of the $2D$ discrete complex-valued distributions ($N \times M$ matrices) at the object and sensor planes, respectively. $A_l \in \mathbb{C}^{n\times n}$ is a discrete forward propagation operator, $n = N \cdot M$. We consider the paraxial approximation of the wave field propagation defined by the Rayleigh-Sommerfield integral. Depending of the used discretization model of this integral, the operators A_l in (1) can be e.g. angular spectrum decomposition [5] or the discrete diffraction transform in the matrix (M-DDT, [6]) or the Fourier transform domains (F-DDT, [7]). In our numerical experiments, we use DDT models enabling the exact pixel-to-pixel mapping of $\mathbf{u}_0$ to $\mathbf{u}_{z_l}$.

According to the used vector-matrix notation, the observation model with the additive Gaussian noise at the sensor planes takes the form:

$$\mathbf{o}_l = |\mathbf{u}_{z_l}|^2 + \boldsymbol{\varepsilon}_l, l = 1, ...L \tag{2}$$

Here we assume for simplicity that the resulting noise is zero-mean Gaussian $\varepsilon_l[k] \sim \mathcal{N}(0, \sigma_l^2)$.

Let us assume that the object amplitude $\mathbf{a}_0 \in \mathbb{R}^n$ and phase $\boldsymbol{\varphi}_0 \in \mathbb{R}^n$ can be separately approximated by a small numbers of non-zero elements: $\boldsymbol{\theta}_a$ in a basis $\boldsymbol{\Psi}_a$ for the object amplitude and $\boldsymbol{\theta}_\varphi$ in a basis $\boldsymbol{\Psi}_\varphi$ for the object phase. $\mathbf{u}_0 = \mathbf{a}_0 \circ exp(j \cdot \boldsymbol{\varphi}_0)$ and "$\circ$" denotes the Hadamard product. Thus, the object wave field is reconstructed from the noisy intensity data $\mathbf{o}_l$, and the object amplitude and phase are processed via sparse decomposing in fixed data dependent bases. It is found that, in contrast to classical orthonormal bases, overcomplete frame based modeling is a much more efficient for imaging [8] and results in a better wave field reconstruction accuracy.

3 Sparse modeling of object amplitude and phase

The sparse approximation of the object amplitude and phase can be given in the synthesis form as $\mathbf{a}_0 = \boldsymbol{\Psi}_a \cdot \boldsymbol{\theta}_a$, $\boldsymbol{\varphi}_0 = \boldsymbol{\Psi}_\varphi \cdot \boldsymbol{\theta}_\varphi$ or in the analysis form as $\boldsymbol{\theta}_a = \boldsymbol{\Phi}_a \cdot \mathbf{a}_0$, $\boldsymbol{\theta}_\varphi = \boldsymbol{\Phi}_\varphi \cdot \boldsymbol{\varphi}_0$. These priori unknown bases are selected from given sets of potential ones, and the vectors $\boldsymbol{\theta}_a$, $\boldsymbol{\theta}_\varphi \in \mathbb{R}^m$ can be considered as spectra $(m \gg n)$ in a parametric data adaptive approximation. The sparsity of approximation is characterized by either the l_0 norm $||\boldsymbol{\theta}||_0$ defined as a number of non-zero components of the vector $\boldsymbol{\theta}$ or the l_1 norm $||\boldsymbol{\theta}||_1 = \sum_s |\boldsymbol{\theta}_s|$. In this work we use l_1 norm recalling that results obtained by l_0 or l_1 norms are shown to be closed to each other [9].

In this paper we apply the newly developed BM3D-frames [4] for a sparse modeling of both the object amplitude and phase: we achieve the reduction of the BM3D-frame domain by thresholding as a solution of the optimization problem.

4 Sparse splitting augmented Lagrangian (SSAL) algorithm

According to the maximum likelihood approach, the reconstruction of the whole wave field is performed by minimization of the following criterion:

$$J = \sum_{l=1}^{L} \frac{1}{2\sigma_l^2} ||\mathbf{o}_l - |\mathbf{u}_{z_l}|^2||_2^2 + \tau_a \cdot ||\boldsymbol{\theta}_a||_{l_p} + \tau_\varphi \cdot ||\boldsymbol{\theta}_\varphi||_{l_p}, \tag{3}$$

subject to (1), $\mathbf{a}_0 = \boldsymbol{\Psi}_a \cdot \boldsymbol{\theta}_a$, $\boldsymbol{\varphi}_0 = \boldsymbol{\Psi}_\varphi \cdot \boldsymbol{\theta}_\varphi$, $\boldsymbol{\theta}_a = \boldsymbol{\Phi}_a \cdot \mathbf{a}_0$, $\boldsymbol{\theta}_\varphi = \boldsymbol{\Phi}_\varphi \cdot \boldsymbol{\varphi}_0$.

The quadratic fidelity term in (3) appears due to our assumption that the observation noise is Gaussian. The following two terms define the sparse regularization in the spectral domain, where the positive parameters τ_a and τ_φ define a balance between the fit of observations, the smoothness of the wave field reconstruction and the complexity of the used model. In our work we are looking for a solution of (3) using the augmented Lagrangian technique as in [10]. Let us replace the constrained minimization by an unconstrained one changing the constraints for sparse modeling by the quadratic penalties with positive weights. Then, the criterion (3) can be divided as $J(\mathbf{u}_0, \{\mathbf{u}_{z_l}\}, \{\mathbf{\Lambda}_l\}, \boldsymbol{\theta}_a, \boldsymbol{\theta}_\varphi) = J_1(\mathbf{u}_0, \{\mathbf{u}_{z_l}\}, \{\mathbf{\Lambda}_l\}, \mathbf{v}_0) + J_2(\mathbf{u}_0, \boldsymbol{\theta}_a, \boldsymbol{\theta}_\varphi)$, where

$$\begin{aligned} J_1 &= \sum_{l=1}^{L} \frac{1}{\sigma_l^2}[\frac{1}{2}||\mathbf{o}_l - |\mathbf{u}_{z_l}|^2||_2^2 + \frac{1}{\gamma_l}||\mathbf{u}_{z_l} - A_l \cdot \mathbf{u}_0||_2^2 + \\ &+\frac{2}{\gamma_l}\operatorname{Re}\{\mathbf{\Lambda}_l^H \cdot (\mathbf{u}_{z_l} - A_l \cdot \mathbf{u}_0)\} + \frac{1}{\xi}||\mathbf{u}_0 - \mathbf{v}_0||_2^2 \end{aligned} \tag{4}$$

$$J_2 = \tau_a \cdot ||\boldsymbol{\theta}_a||_{l_p} + \tau_\varphi \cdot ||\boldsymbol{\theta}_\varphi||_{l_p} + \frac{1}{2\gamma_a}||\boldsymbol{\theta}_a - \mathbf{\Phi}_a \cdot \mathbf{a}_0||_2^2 + \frac{1}{2\gamma_\varphi}||\boldsymbol{\theta}_\varphi - \mathbf{\Phi}_\varphi \cdot \boldsymbol{\varphi}_0||_2^2 \tag{5}$$

The synthesis constraints in (4) are given in the complex-valued form as the approximation of $\mathbf{u}_0$ by $\mathbf{v}_0 = \mathbf{\Psi}_a \cdot \boldsymbol{\theta}_a \circ exp(j \cdot \mathbf{\Psi}_\varphi \cdot \boldsymbol{\theta}_\varphi)$. In Eq. (5) γ_a and γ_φ are positive penalty parameters introduced for the analysis constraints and ξ is the penalty parameter for the synthesis constraint.

A typical Lagrangian based optimization assumes minimization of the criterion J with respect to $\mathbf{u}_0$, $\{\mathbf{u}_{z_l}\}$, $\boldsymbol{\theta}_a, \boldsymbol{\theta}_\varphi$ and its maximization on the vectors of the Lagrange multipliers $\{\mathbf{\Lambda}_l\} \in \mathbb{C}^{n\times 1}$. $(\cdot)^H$ stands in Eq. (4) for the Hermitian conjugate, and γ_l are positive penalty coefficients. Instead of optimization of $J_1 + J_2$ we use a partial alternative minimization of these two summands: J_1 on $\mathbf{u}_0$, performing inversion of the forward wave field propagation, and J_2 on $\boldsymbol{\theta}_a, \boldsymbol{\theta}_\varphi$ performing filtering of the object amplitude and phase in a spectral domain. The splitting variable $\mathbf{v}_0$ separates minimizations on $\mathbf{u}_0$ and $\boldsymbol{\theta}_a, \boldsymbol{\theta}_\varphi$. Since the minimization of J_1 on $\mathbf{u}_0$ in general results in increasing of J_2, and vice versa, we are looking for a fixed-point $(\mathbf{u}_0^*, \boldsymbol{\theta}_a^*, \boldsymbol{\theta}_\varphi^*)$ as a compromise in such a selfish behavior. Then, a complex-valued object wave field is reconstructed using the inverse imaging algorithm with decoupling of inversion and filtering of both the amplitude and phase [4]:

$$\begin{aligned} (\boldsymbol{\theta}_a^*, \boldsymbol{\theta}_\varphi^*) &= \arg\min_{\boldsymbol{\theta}_a, \boldsymbol{\theta}_\varphi} J_2(\mathbf{a}_0^*, \boldsymbol{\varphi}_0^*, \boldsymbol{\theta}_a, \boldsymbol{\theta}_\varphi) \\ \mathbf{v}_0^* &= \mathbf{\Psi}_a \cdot \boldsymbol{\theta}_a^* \circ exp(j \cdot \mathbf{\Psi}_\varphi \cdot \boldsymbol{\theta}_\varphi^*) \\ \mathbf{u}_0^* &= \arg\min_{\mathbf{u}_0, \{\mathbf{u}_{z_l}\}} \max_{\{\mathbf{\Lambda}_l\}} J_1(\mathbf{u}_0, \{\mathbf{u}_{z_l}\}, \{\mathbf{\Lambda}_l\}, \mathbf{v}_0^*) \end{aligned} \tag{6}$$

In [10] we show that the object wave field estimate can be easily achieved if the optimization variables (here $\mathbf{u}_0, \{\mathbf{u}_{z_l}\}, \{\mathbf{\Lambda}_l\}$) are partioned into several blocks according to their role. Then, the resulting augmented Lagrangian function J_1 is minimized with respect to each block by fixing all other blocks at each inner iteration. The criterion J_2 can be divided into two parts with respect to the object amplitude and phase. From the minimum condition for J_1 and J_2 in (6) we arrive at the proposed iterative algorithm.

0. Initialization for $t = 0$: $\mathbf{u}_0^0 = \mathbf{a}_0^0 \circ exp(j \cdot \boldsymbol{\varphi}_0^0)$, $\{\mathbf{\Lambda}_r^0\}$, $\mathbf{\Phi}_a$, $\mathbf{\Phi}_\varphi$, $\mathbf{\Psi}_a$, $\mathbf{\Psi}_\varphi$

Repeat for $t = 1, 2...$ (7)

1. $\boldsymbol{\theta}_a^t = \Im h_{\tau_a \gamma_a}(\boldsymbol{\Phi}_a \cdot \mathbf{a}_0^{t-1})$, $\boldsymbol{\theta}_\varphi^t = \Im h_{\tau_\varphi \gamma_\varphi}(\boldsymbol{\Phi}_\varphi \cdot \boldsymbol{\varphi}_0^{t-1})$
2. $\mathbf{v}_0^t = \boldsymbol{\Psi}_a \cdot \boldsymbol{\theta}_a^t \circ exp(j \cdot \boldsymbol{\Psi}_\varphi \cdot \boldsymbol{\theta}_\varphi^t)$
3. $\mathbf{u}_{z_l}^{t+1/2} = A_l \cdot \mathbf{u}_0^t,\ l = 1, ...L$
4. $\mathbf{u}_{z_l}^{t+1}[k] = \mathcal{G}(\mathbf{o}_r[k], \mathbf{u}_{z_l}^{t+1/2}[k], \boldsymbol{\Lambda}_l^t[k]),\ l = 1, ...L$
5. $\boldsymbol{\Lambda}_l^{t+1} = \boldsymbol{\Lambda}_l^t + \alpha_l \cdot (\mathbf{u}_{z_l}^{t+1} - \mathbf{u}_{z_l}^{t+1/2}),\ ,l = 1, ...L$
6. $\mathbf{u}_0^{t+1} = (\sum_{l=1}^{L} \frac{1}{\gamma_l \sigma_l^2} A_l^H A_l + \frac{1}{\xi} \cdot \mathbf{I}_{n \times n})^{-1} \times (\sum_{l=1}^{L} \frac{1}{\gamma_l \sigma_l^2} A_l^H \cdot [\mathbf{u}_{z_l}^{t+1} + \boldsymbol{\Lambda}_l^t] + \frac{\mathbf{v}_0^t}{\xi})$,

End on t

The initialization for $t = 0$ concerns the object distribution (e.g. $\mathbf{u}_0^0[k] = 1/2$), Lagrangian multipliers (e.g. $\{\boldsymbol{\Lambda}_r^0[k]\} = 0$), and the BM3D-frame based bases for synthesis and analysis for both the object amplitude and phase.

The updates of $\{\mathbf{u}_{z_l}^{t+1}\}$ is realized by fitting of $\{\mathbf{u}_{z_l}^{t+1/2}\}$ to the observations $\mathbf{o}_r$ by the operator $\mathcal{G} = \arg\min_{\{\mathbf{u}_{z_l}\}} J_1$ (see the definition in [10]). We call this procedure of optimization the sparse splitting augmented Lagrangian ($SSAL$). The main difference with the original AL algorithm originated in [10] is that the wave field estimates at the object and sensor planes are calculated using filtering in a BM3D spectral domain. The used algorithm of BM3D filter can be divided into three steps:

1. *Analysis*. Highly correlated image blocks are distinguish and stacked together to form a 3D data array, which is decorrelated by an invertible 3D transform (calculation of spectra $\boldsymbol{\theta}_a^t$, $\boldsymbol{\theta}_\varphi^t$).

2. *Processing*. 3D group spectra obtained from 3D data array separately for the object amplitude and phase are filtered by thresholding (the result of Step 1 in the proposed $SSAL$ algorithm).

3. *Synthesis*. Filtered spectra are inverted providing estimates for each block in a group. These blocks are returned to their original positions, then the final image estimate is aggregated by weighted averaging over all block-wise estimates (Step 2).

It is easy to see that depending on the chosen l_1 or l_0 norms we use '*soft*' or '*hard*' thresolding denoted as [4], [8]:

$$\boldsymbol{\theta} = \Im h_\tau(\mathbf{u}) = \{ \begin{array}{c} sign(\mathbf{u}) \cdot (|\mathbf{u}| - \tau)_+,\ \text{if } l_p = l_1 \\ \mathbf{u} \circ \mathbf{I}(|\mathbf{u}| > \sqrt{2\tau}),\ \text{if } l_p = l_0 \end{array} \tag{8}$$

For different norms l_p in Eq. (8) we consider the optimization problem in the form $\frac{1}{2}||\boldsymbol{\theta} - \mathbf{u}||_2^2 + \tau \cdot ||\boldsymbol{\theta}||_{l_p} \rightarrow \min$.

Taking into account the additive nature of the norms this problem can be solved independently for each component of $\boldsymbol{\theta}_i$, $i = 1, ...m$. If $l_p = l_1$ the minimum condition leads to $|\boldsymbol{\theta}_i| = |\mathbf{u}_i| + \tau \cdot sign(\boldsymbol{\theta}_i)$.

The solution for $l_p = l_0$ can be easily found using $|\boldsymbol{\theta}_i|_0 = 1(\boldsymbol{\theta}_i \neq 0)$. The resulting non-convex function $\frac{1}{2}|\boldsymbol{\theta}_i - \mathbf{u}_i|^2 + \tau \cdot 1(\boldsymbol{\theta}_i \neq 0)$ has two minima equal to τ in $\boldsymbol{\theta}_i = 0$ and $\boldsymbol{\theta}_i = \mathbf{u}_i$ if $\mathbf{u}_i = \sqrt{2\tau}$. For all $\mathbf{u}_i < \sqrt{2\tau}$ the minimum of this function is in $\boldsymbol{\theta}_i = 0$.

Note that the threshold in the algorithm (7) is $\tau_a\gamma_a$ for the BM3D spectrum of the object amplitude and $\tau_\varphi\gamma_\varphi$ for the phase spectrum.

5 Numerical experiments

In our simulation experiments, we compare three algorithms: the recent AL method from [10], the successive $SBMIR$ from [3] and the proposed $SSAL$ algorithm (7). Here we consider a phase-only object distributions given as $\mathbf{u}_0 = 1 \cdot exp(j \cdot \pi(\mathbf{w} - 1/2))$, where $0.1 \leq \mathbf{w} \leq 1$ is the binary test-images *chessboard* (128×128). $\{\mathbf{u}_{z_l}\}$ and $\mathbf{u}_0$ are pixilated with square pixels $\Delta \times \Delta$, $\Delta = 6.7\mu m$ with 100% fill factors. The presented results are given for $L = 5$ noisy observations with $\sigma_l = \sigma = 0.05$ for all l, and the following setup parameters: wavelength $\lambda = 532nm$, $\Delta_z = 2mm$, $z_1 = 2 \cdot z_f$, z_f is "in-focus" distance [11]. Note that the crucial point of the $SSAL$ wave field reconstruction is its initial guess for $\mathbf{u}_0^0$. We are looking for this initial object estimate by AL. It is found that the best reconstruction accuracy can be achieved for a compromise between overall sharpness of the object reconstruction and smoothness

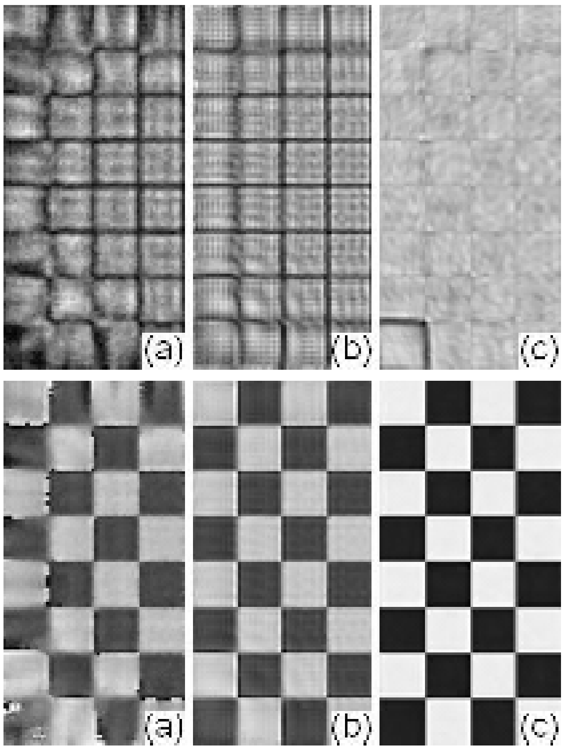

Figure 1: Fragments of the reconstructed (top image) amplitude and (bottom) phase, obtained by (a) $SBMIR$, $RMSE(\mathbf{a}_0)$=0.35, $RMSE(\boldsymbol{\varphi}_0)$=0.58; (b) AL, $RMSE(\mathbf{a}_0)$=0.23, $RMSE(\boldsymbol{\varphi}_0)$=0.26 and (c) $SSAL$, $RMSE(\mathbf{a}_0)$=0.026, $RMSE(\boldsymbol{\varphi}_0)$=0.036.

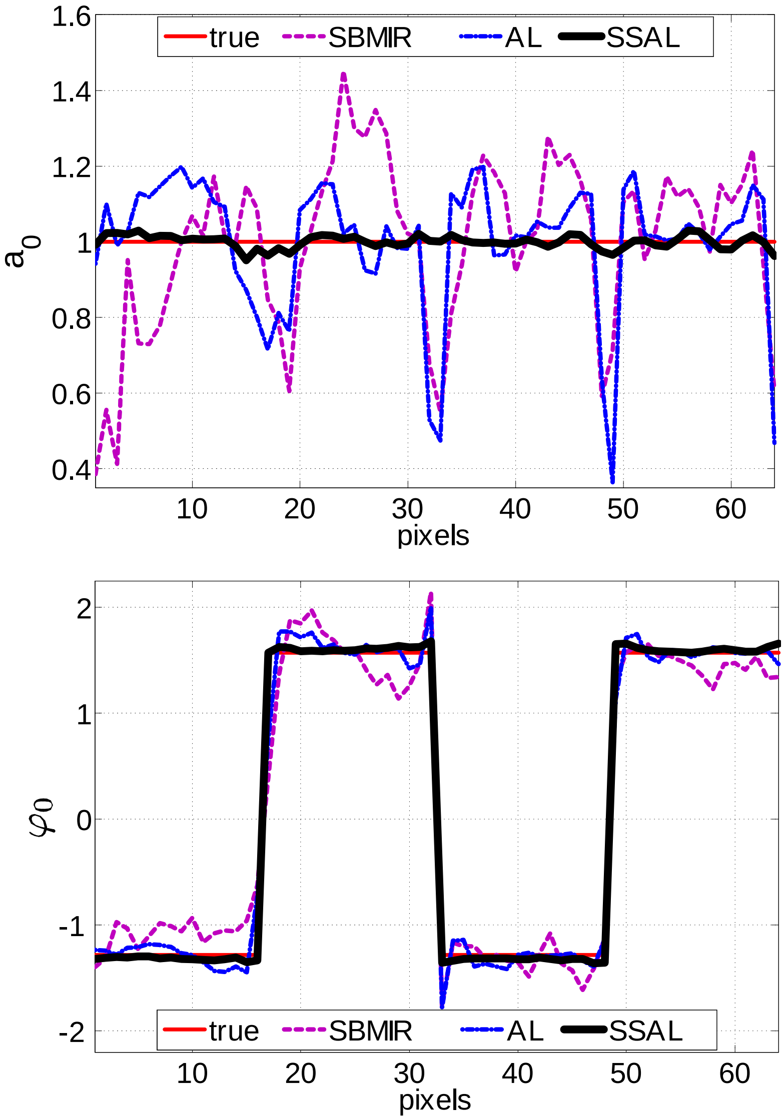

Figure 2: Cross-sections of the reconstructed (top image) object amplitude and (bottom) phase.

of its details. Thus, here we use 50 iterations for the object initialization by AL and 50 iterations of the $SSAL$ algorithm.

In Fig. 1, the reconstructed object amplitude and phase are shown after 100 iterations of the considering phase-retrieval algorithms. The visual advantage of the proposed $SSAL$ algorithm is obvious. The reconstruction accuracy is given in root-mean-square error ($RMSE$) values for the whole image. The corresponding cross-sections are illustrated in Fig. 2 with the best performance obtained by $SSAL$: this results are very close to the true value, while the AL and $SBMIR$ reconstructions are blurred and have a quite large deviation.

Numerical experiments demonstrate a significant better reconstruction quality (via $RMSE$) of $SSAL$: here it is approximately ten times better with respect to AL and more for $SBMIR$.

6 Conclusions

The proposed ML-like phase-retrieval technique takes into account the Gaussian noise distribution and prior information on the object: the developed algorithm can be treated as a further development of the recent *AL* [10]. It is shown, that the object sparse regularization via BM3D-frame based filtering dramatically improves the reconstruction accuracy and imaging.

The Matlab code used for numerical simulations and more materials are available on our web page http://www.cs.tut.fi/~lasip/DDT/

References

[1] J. R. Fienup, "Phase retrieval algorithms: A comparison," *Appl. Opt.* 21, 2758-2769 (1982).

[2] D. L. Misell, "A method for the solution of the phase problem in electron microscopy," *J. Phys. D 6*, L6–L9 (1973).

[3] 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).

[4] V. Katkovnik, A. Danielyan and K. Egiazarian, "Decoupled inverse and denoising for image deblurring: variational BM3D-frame technique," *ICIP-2011* (2011), submitted.

[5] J. W. Goodman, *Introduction to Fourier Optics*, 2nd. ed., New York: McGraw-Hill Inc (1996).

[6] 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, pp. 3407-3423 (2009).

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

[8] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, Springer (2010).

[9] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory* 52, 1289–1306 (2006).

[10] A. Migukin, V. Katkovnik and J. Astola, "Wave field reconstruction from multiple plane intensity-only data: augmented Lagrangian algorithm," JOSAA 28, (2011) in press.

[11] Th. Kreis, *Handbook of Holographic Interferometry, Optical and Digital Methods*, Wiley-VCH, (2005).