

Phase retrieval from noisy data based on sparse approximation of object phase and amplitude

Vladimir Katkovnik

Abstract

A variational approach to object phase and amplitude reconstruction from multiple Poissonian intensity observations is developed for the typical phase retrieval scenario. Sparse modeling of amplitude and absolute phase of the object is one of the key elements of the derived SPAR algorithm. We introduce also a simplified version of this algorithm where the sparse modeling of object phase and amplitude is omitted (GS algorithm). The efficiency of these algorithms is demonstrated by simulation experiments for the coded diffraction pattern scenario. The comparison is produced versus the truncation Wirtingler flow (TWF) algorithm (Y. Chen and E. J. Candès, 2015, <http://statweb.stanford.edu/~candes/papers/TruncatedWF.pdf>), which is the state-of-the-art in the field. For noisy observations SPAR demonstrates a definite advantage over TWF. For low noise level the accuracy of SPAR, GS and TWF is nearly identical. The GS algorithm is about twice faster than TWF, while SPAR computationally much more demanding is slower than TWF and GS.

Index Terms

Complex domain imaging, phase retrieval, sparse complex domain signal reconstruction.

I. INTRODUCTION

A. Phase retrieval formulation

Vladimir Katkovnik with Department of Signal Processing, Technology University of Tampere, Tampere, Finland. E-mail: vladimir.katkovnik@tut.fi.

This work is supported by Academy of Finland, project no. 287150, 2015-2019.

TRANSSPARENT specimens, for instance biological cells, are nearly transparent with respect to intensity of passing light. However, variations in thickness, density and refractive indices of studied objects result in corresponding variations of phase delays. *Visualization* of these phase variations by transforming them in light intensity is one of the challenging problems in optics. The revolutionary *phase contrast imaging* (Frits Zernike 1930s, Nobel prize 1953) solves the problem by introducing a modulation of the wavefront in the focal (Fourier) plane of the principal lens. In this way *the desirable visualization* is achieved where the observed light intensity is linked with phase variations. However, this visualization is only qualitative because there is no one-to-one relations between the observed phase and variations of object properties. Nevertheless the phase contrast microscopy is one of the most frequently applied optical methods in many areas of research and applications.

Quantitative phase visualization is targeted on precise phase imaging. In the modern development it is fundamentally based on digital data processing. One of the popular techniques of the quantitative phase visualization is *phase retrieval*.

The phase retrieval from multiple experiments is formulated as finding a complex-valued vector $\mathbf{x} \in \mathbb{C}^n$ from a set of real-valued observations $\mathbf{y}_s \in \mathbb{R}^m$:

$$\mathbf{y}_s = |\mathbf{A}_s \mathbf{x}|^2, \quad s = 1, \dots, L. \quad (1)$$

In terms of the coherent diffractive imaging the model (1) allows the following interpretation. Provided the unit intensity of a laser beam the vector \mathbf{x} is an object (specimen) transfer function and the complex-valued wavefront just behind the object (see Fig.1a); $\mathbf{A}_s \in \mathbb{C}^{m \times n}$ is an $m \times n$ matrix of the wavefront propagation operator from the object to the sensor plane and the vector \mathbf{y}_s is an intensity of the wavefront registered by the sensor. The squared absolute value in (1) is an element-wise operation. Thus, the items of the vector \mathbf{y}_s are squared absolute values of the corresponding items of the vector $\mathbf{A}_s \mathbf{x} \in \mathbb{C}^m$.

Assume for a moment that the complex-valued $\mathbf{A}_s \mathbf{x}$ denoted as \mathbf{u}_s are known then the quadratic equations (1) can be replaced by the linear ones $\mathbf{A}_s \mathbf{x} = \mathbf{u}_s$ and finding of \mathbf{x} is reduced to the linear algebra problem. Thus, the phases of $\mathbf{A}_s \mathbf{x}$ eliminated by the modulus in (1) transform a simple linear problem in a much more complex quadratic one.

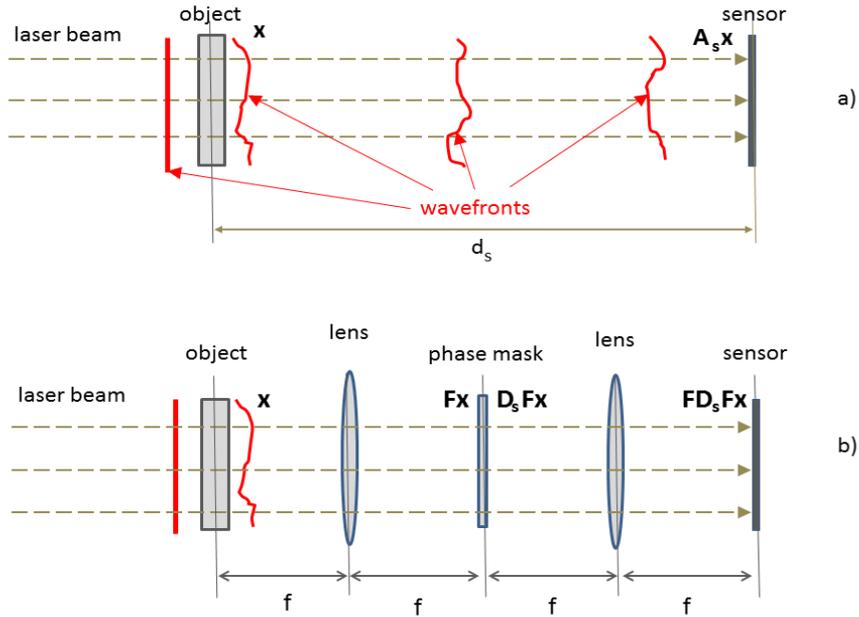


Fig. 1. Examples of optical setups for lensless (a) and $4f$ lens (b) scenarios for phase retrieval.

Conventionally, the term *phase retrieval* is addressed to reconstruction of the missing phases of the vectors $\mathbf{A}_s \mathbf{x}$. However, in (1) the phase of the complex-valued object $\mathbf{x} \in \mathbb{C}^n$ is also unknown and actually reconstruction (retrieval) of this object phase is the main goal of the problem at hand.

In this paper we refocus the standard setting of the phase retrieval problem by treating the missed phases of $\mathbf{A}_s \mathbf{x}$ as auxiliary variables and phase and amplitude of the object \mathbf{x} as the main unknowns of interest. We show that the sparse modeling applied to \mathbf{x} (to both phase and amplitude) leads to an efficient algorithm solving the problem.

In this interpretation the phase retrieval becomes the quantitative phase imaging problem with the real-valued observations $\mathbf{y}_s \in \mathbb{R}^m$ given by (1).

2D imaging is studied in this paper. The vectors and matrixes in the above equations correspond to the vectorized representations conventionally applied for this kind of the 2D problems.

Design of the image formation operators \mathbf{A}_s in (1) is a crucial moment in order to gain observation diversity sufficient for reliable reconstruction of object phase as well as object amplitude. Pragmatically, this diversity means that the set $\{\mathbf{A}_s\}_1^L$ consists of the operators so different that the observations \mathbf{y}_s suffice finding \mathbf{x} . Image defocussing is one of the popular ways to get this kind of diverse observations. First results based on the defocus approach

are demonstrated in [1] and [2] for two defocussed images obtain using optical lens. A joint estimation of object and aberrations in incoherent imaging system from multiple images incorporating phase diversity by defocussing is studied in [3].

In the coherent lensless imaging (Fig.1a) the laser beam goes through the object and after free-space propagation the intensity of the diffracted wavefield is registered by the sensor array. The corresponding operators \mathbf{A}_s depends on the object-sensor distance d_s and the wavelength λ_s , $\mathbf{A}_s = \mathcal{A}(d_s, \lambda_s)$, where the operator \mathcal{A} can be modelled by the rigorous Rayleigh-Sommerfeld integral or by the Fresnel and Fraunhofer approximations of this integral [4]. The latter modeling is equivalent to the discrete Fourier transform \mathbf{F}_s for the operator \mathbf{A}_s because the phase factor of the Franhofer transform is cancelled by the modulus operation in (1).

The approach using experiments with a set of distances d_s (displaced sensor planes) is studied in [5], [6] and [7]. In the recent development, in particular we refer to [8] and [9], a spatial light modulator (SLM) is exploited in order to get a set of differently defocussed images.

A phase modulation of the wavefront near the object plane is another popular tool to get diverse observations. The phase modulation at the object plane plus the Fraunhofer wavefront propagation result in the observation model known as a coded diffraction pattern [10]:

$$\mathbf{y}_s = |\mathbf{F}\mathbf{D}_s\mathbf{x}|^2, s = 1, \dots, L, \quad (2)$$

where $\mathbf{F} \in \mathbb{C}^{n \times n}$ denotes the Fourier transform and $\mathbf{D}_s \in \mathbb{C}^{n \times n}$ is a diagonal matrix of complex exponents, $\mathbf{D}_s = \text{diag}\{\exp(j\phi_1(s)), \exp(j\phi_2(s)), \dots, \exp(j\phi_n(s))\}$.

This phase modulation can be implemented by a special phase mask inserted just behind the object plane in Fig.1a.

The phases $\phi_k(s)$ in \mathbf{D}_s can be generated as random. Let $\phi_k(s)$ be i.i.d. zero mean Gaussian, $\phi_k(s) \sim N(0, \sigma)$. Then, $d_{ks} = \exp(j\phi_k(s))$ are random elements of \mathbf{D}_s , such that $E\{d_{ks}\} = 0$ and the matrices $\mathbf{A}_s = \mathbf{F}\mathbf{D}_s$ are zero-mean random. This random phase modulation changes the spectrum of $\mathbf{F}\mathbf{x}$ in a radical way extending a distribution of the intensity from low to high frequency components.

The phase modulation (coded aperture imaging) is applied in various optical setups. For instance, consider the coherent two lens $4f$ optical system shown in Fig.1b, where f is a focal length of the lenses. The wavefront in the

focal plane behind the first lens is the Fourier transform of the wavefront going through the object [4]. Further, the wavefront at the sensor plane, which locates at the focal length of the second lens, is the Fourier transform of the wavefront in the Fourier plane of the first lens. The phase modulation is produced using phase mask or SLM in the focal (Fourier) plane of the first lens. The intensity registered by the sensor in this system can be represented as

$$\mathbf{y}_s = |\mathbf{F}\mathbf{D}_s\mathbf{F}\mathbf{x}|^2, \quad s = 1, \dots, L. \quad (3)$$

It is demonstrated in [11] that the phases $\phi_k(s)$ in \mathbf{D}_s can be selected in such way that the distributions of the wavefront at the sensor plane imitates desired displacements $\{d_s\}$ of the sensor plane with respect to the object plane. The phase retrieval for phase imaging in the $4f$ optical setup with the random phase modulation in the Fourier plane is studied in [12].

Developments of the phase modulation techniques with various applications can be seen in the book [13].

B. Phase retrieval algorithms

Let us start from the popular Gerchberg-Saxton (GS) algorithms (e.g. [14], [15]). These iterative algorithms are based on alternating projections between the object plane with a complex-valued \mathbf{x} and the Fourier (diffraction) plane $\mathbf{A}\mathbf{x}$ with a given (measured) amplitude \mathbf{z} . At the Fourier plane the amplitudes in the vectors $\mathbf{A}\mathbf{x}$ are replaced by square roots of the corresponding items of \mathbf{z} . The back projection of this result to the object plane is modified according to the prior information on the object, e.g. support size and shape, amplitude value of \mathbf{x} , etc. The GS algorithms exist in various modifications. In particular, the phase retrieval based on the maximum likelihood technique for Gaussian and Poissonian observations is developed in [3] for phase diversity enabled by multiple defocussing of observations. The review and analysis of the GS algorithms as well as further developments can be seen in the recent paper [16].

The algorithms, known as a single-beam multiple-intensity reconstruction (SBMIR), are targeted on reconstruction of $3D$ wavefield covered by $2D$ intensity measurement planes (e.g. [5], [6], [7]). "The SBMIR algorithm starts from an initial guess of the wavefront at the first measurement plane. Then, this initial guess propagates numerically forward from the one measurement plane to the next following one successively through the all sequence of the

measurements. At each plane the calculated modulus of the wave field is replaced by the square root of the intensity measured for this plane according to the GS algorithm. When the last measurement plane is reached the wave field estimate at this plane is propagated back to the first plane. This iterative process is repeated until convergence.”

Contrary to the intuitively clear heuristic GS algorithms the variational approaches to phase retrieval usually have a stronger mathematical background including image formation modeling, formulation of the objective function (criterion) and finally going to numerical techniques solving corresponding optimization tasks. Here we wish refer to the recent overview [17] concentrated on the algorithms for the phase retrieval models with Fourier transform measurements of the form $y_s = |\mathbf{F}_s \mathbf{x}|^2$. The constraints sufficient for uniqueness of the solution are presented in detail. Beyond the alternating projection GS a few novel mathematical methods are discussed: semidefinite programming phase lifting using matrix completion (*PhaseLift algorithm*) [18] and greedy sparse phase retrieval (*GESPAR algorithm*) [19]. Some optical applications of the phase retrieval algorithms are considered in the overview paper [17].

Many publications concern variational techniques as well as revisions of the intuitive GS algorithms by using optimization formulations. In particular, the links between the conventional GS and variational techniques are studied in [20] and [21]. A variational formulation for the phase retrieval is demonstrated in [22], where the criterion corresponding to Poissonian observations and the prior defining the smoothness of the phase are proposed. The problem is formalized as a penalized likelihood optimization. The conjugate gradient iterative algorithm for this setting is proposed in [23].

Especially we wish to note the recent Wirtingler flow (WF) algorithms presented in [24] and [25]. These algorithms are iterative complex domain gradient descents applied to Poissonian likelihood criteria. Specific features of these algorithms are as follows: a special spectral initialization, a non-trivial growing step-size parameter and truncation of the gradient in the truncation Wirtingler flow (TWF) version of the algorithms [24]. Meticulous mathematical analysis is produced for the algorithm design, parameter selection and performance evaluation. It is stated that the solution of the quadratic equations (1) can be done “*nearly as easy as solving linear equation*”. In this mathematical analysis the elements of the matrices \mathbf{A}_s in (1) are random independent and subject of a complex-valued Gaussian distribution. Simulation experiments demonstrate that the TWF algorithm works and works very

well provided a small level of the random noise.

C. Contribution and structure of this paper

The sparsity based techniques is a hot topic in phase retrieval. The signal domain sparsity is exploited in the GESPAR algorithm [19]: the length of the vector-solution \mathbf{x} is minimized. The transform domain sparsity for amplitude and phase of \mathbf{x} is a base of the Sparse Phase Amplitude Reconstruction (SPAR) algorithm [12]. This approach has been demonstrated for high-accuracy phase imaging in various setups [26], [27], [28].

In this paper this SPAR technique is developed for the phase retrieval problem with the intensities defined in the form (1) and for Poissonian observations.

The SPAR phase retrieval algorithm is derived from the variational formulation of the problem and incorporates two types of filtering: filtering of Poissonian noise at the sensor plane and filtering of phase and amplitude at the object plane. If both these filters are omitted the SPAR algorithm becomes quite similar to the conventional GS algorithm. Hereafter we use the term *GS algorithm* for this *simplified version* of the SPAR algorithm.

Surprisingly, this novel GS algorithm demonstrates the performance nearly identical to the advanced TWF algorithm [25]. Both algorithms enable similar accuracy for the phase and amplitude reconstruction as well as the similar computational complexity.

The complete SPAR algorithm is computationally more demanding than the GS algorithm. For noisy data it demonstrates much higher accuracy as compared with respect to both TWF and GS algorithms. The phase unwrapping is included in the iterations of the SPAR algorithm, when the phase variation overcomes 2π range. It allows to achieve an efficient noise suppression and accurate absolute phase reconstruction.

The paper is organized as follows. In Section II we consider different variational formulations of phase retrieval including derivation of the new GS algorithm and the sparsity modelling for phase and amplitude. The complete SPAR algorithm development is a subject of Section III, where step-by-step solutions of the variational problems are discussed and the SPAR algorithm is composed. Section IV concerns the experimental study of the proposed GS and SPAR algorithms and their comparison versus the TWF algorithm.

II. PROBLEM FORMULATION

A. Sparse wavefront modeling

It is recognized that many natural images (and signals) admit sparse representations, i.e. they can be well approximated by linear combinations of a small number of functions. This is a consequence of the so-called patch-wise self-similarity of these images. It means that it is possible to find in them many quite similar small size patches located in different parts of the image. The sparse and redundant representations is of a special interest in the last years. This interest follows from importance that low dimensional models play in many signal and image applications such as compression, restoration, classification, just to mention a few of them [30].

Let $\mathbf{x} \in \mathbb{C}^n$ be a complex-valued wavefront. Denote $\mathbf{b} = \text{abs}(\mathbf{x})$ and $\varphi = \text{angle}(\mathbf{x}) \in [-\pi, \pi)$ as, respectively, the corresponding images of amplitude (modulus) and the wrapped phase, φ . Then we have $\mathbf{x} = \mathbf{b} \circ \exp(j\varphi)$, where 'o' stands for the element-wise (Hadamard) product of two vectors. Herein, all functions applied to vectors are to be understood as component-wise.

With the intention of formulating treatable phase imaging problems, most approaches follow a two-step procedure: in the first step, an estimate of the so-called principal (wrapped, interferometric) phase in the interval $[-\pi, \pi)$ is determined; in the second step, termed phase unwrapping, the absolute phase is reconstructed by adding of an integer number of 2π multiples to the estimated interferometric phase [31]. In what follows, we denote the principal phase as φ and the absolute phase as φ_{abs} . We introduce the phase-wrap operator $\mathcal{W} : \mathbb{R} \mapsto [-\pi, \pi)$, linking the absolute and principal phase as $\varphi = \mathcal{W}(\varphi_{abs})$. We also define the unwrapped phase as $\varphi_{abs} = \mathcal{W}^{-1}(\varphi)$. Notice that \mathcal{W}^{-1} is not an inverse operator for \mathcal{W} because the latter is highly non-linear and for signals of dimension two and higher there is no one-to-one relation between φ_{abs} and φ .

In sparse coding for complex valued \mathbf{x} , we may think in two different directions: either we use a complex-valued sparse representation to model directly for the complex image \mathbf{x} , as recently proposed in [32] and [33], or we use separate sparse real-valued representations for the amplitude \mathbf{b} and absolute phase φ_{abs} images of \mathbf{x} . To some extent the choice of the type of the sparse modeling depends on the application.

In this paper, we follow the second type of the wavefront modeling. Following to [12] we use this sparse wavefront

modeling as the following matrix operations:

$$\mathbf{b} = \Psi_a \boldsymbol{\theta}_a, \quad \boldsymbol{\varphi} = \Psi_\varphi \boldsymbol{\theta}_\varphi, \quad (4)$$

$$\boldsymbol{\theta}_a = \Phi_a \mathbf{b}, \quad \boldsymbol{\theta}_\varphi = \Phi_\varphi \boldsymbol{\varphi}_{abs}, \quad (5)$$

where $\boldsymbol{\theta}_a \in \mathbb{R}^p$ and $\boldsymbol{\theta}_\varphi \in \mathbb{R}^p$ are, respectively, the amplitude and absolute phase spectra of the object \mathbf{x} . In (4), the amplitude $\mathbf{b} \in \mathbb{R}^n$ and absolute phase $\boldsymbol{\varphi} \in \mathbb{R}^n$ are synthesized from the amplitude and phase spectra $\boldsymbol{\theta}_a$ and $\boldsymbol{\theta}_\varphi$. On the other hand, the analysis Eqs.(5) give the spectra for amplitude and phase of the wavefront \mathbf{x} . In (4)-(5) the synthesis ($n \times p$) and analysis ($p \times n$) matrices are denoted as Ψ_a , Ψ_φ and Φ_a , Φ_φ , respectively.

Following the sparsity rationale we assume that amplitude and phase spectra, $\boldsymbol{\theta}_a$ and $\boldsymbol{\theta}_\varphi$, respectively, are sparse; i.e., most elements thereof are zero. In order to quantify the level of sparsity of $\boldsymbol{\theta}_a$ and $\boldsymbol{\theta}_\varphi$, i.e., their number of non-zero (active) elements, we use the pseudo l_0 -norm $\|\cdot\|_0$ defined as a number of non-zero elements of the vector-argument. Therefore, we will design estimation criteria promoting low values of $\|\boldsymbol{\theta}_a\|_0$ and $\|\boldsymbol{\theta}_\varphi\|_0$.

Usually, the spectral dimensions are much higher than the dimensions of the image \mathbf{x} , $p \gg n$, while the number of the active elements, i.e. the value of the pseudo l_0 -norms of spectra, are much smaller than p and can be smaller or larger than n . The sparse approximations in the form (4)-(5) are initiated from the works [26] and [27].

It is obvious that for the complex exponent there is no difference between the principal and absolute phase, $\exp(j\boldsymbol{\varphi}_{abs}) = \exp(j\boldsymbol{\varphi})$, and the angle operator $\text{angle}(\mathbf{x})$ gives the principal phase. However, there is a great deal of difference between the sparsity for the absolute and interferometric phases. It is because in many applications the absolute phase can be smooth or piece-wise smooth function easily allowing sparsification while provided $\max(\text{abs}(\boldsymbol{\varphi}_{abs})) > \pi$ the corresponding wrapped phase experiences multiple heavy discontinuities. This kind of images are known as interferometric fringe patterns and they are quite difficult for direct sparse modelling. Nevertheless, an efficient sparsification of the wrapped phase can be achieved through approximation of the complex exponent $\exp(j\boldsymbol{\varphi})$. Here we wish to mention the windowed Fourier transform developed for fringe processing is [34] as well as different forms of the Gabor transform which are definitely good candidates for this problem.

Another style of the data adaptive efficient approximators for the complex exponent are proposed in already mentioned papers [32] and [33] based on the leaning dictionary techniques and high-order SVD non-local complex domain approximations.

B. Noisy observation modeling

The measurement process in optics amounts to count the photons hitting the sensor's elements and is well modeled by independent Poisson random variables: the probability that a random Poissonian variable $\mathbf{z}_s[l]$ of the mean value $\mathbf{y}_s[l]$ takes a given non-negative integer k , is given by

$$p(\mathbf{z}_s[l] = k) = \exp(-\mathbf{y}_s[l]\chi) \frac{(\mathbf{y}_s[l]\chi)^k}{k!}, \quad (6)$$

where $\mathbf{y}_s[l]$ is the intensity of the wavefront at pixel l defined by (1).

The parameter $\chi > 0$ in (6) is a scaling factor, which can be interpreted as an exposure time or as a sensitivity of the sensor. Recall that the mean and the variance of Poisson random variable $\mathbf{z}_s[l]$ are equal and are given by $\mathbf{y}_s[l]\chi$, i.e., $E\{\mathbf{z}_s[l]\} = \text{var}\{\mathbf{z}_s[l]\} = \mathbf{y}_s[l]\chi$. Defining the observation signal-to-noise ratio (SNR) as the ratio between the square of the mean and the variance of $\mathbf{z}_s[l]$, we have $SNR = E^2\{\mathbf{z}_s[l]\}/\text{var}\{\mathbf{z}_s[l]\} = \mathbf{y}_s[l]\chi$. Thus, the relative noisiness of observations becomes stronger as $\chi \rightarrow 0$ ($SNR \rightarrow 0$) and approaches zero when $\chi \rightarrow \infty$ ($SNR \rightarrow \infty$). The latter case corresponds to the noiseless scenario, i.e. $\mathbf{z}_s[l]/\chi \rightarrow \mathbf{y}_s[l]$ with the probability 1.

The scale parameter χ is of importance for modeling as it allows to control a level of noise in observations. For real data processing one usually can take $\chi = 1$ assuming image scaling such that $E\{\mathbf{z}\} = \mathbf{y}$.

III. ALGORITHM DEVELOPMENT

We formulate reconstruction of the wavefront as a variational problem with estimation of the amplitude and phase of $\mathbf{x} = \mathbf{b} \circ \exp(j\boldsymbol{\varphi})$ from noisy Poissonian observations \mathbf{z}_s . This problem is a rather challenging mainly due to nonlinearity of \mathbf{x} with respect to the amplitude and phase and the periodic nature of \mathbf{x} with respect to the phase.

A. GS algorithm

Let us start from a simplified setting of the problem. Assume that the sparsity hypotheses for the amplitude and phase of \mathbf{x} are not imposed. Then, the maximum likelihood concept for the observations (6) gives the criterion

$$\mathcal{L}(\{\mathbf{u}_s\}) = \sum_{s=1}^L \sum_{l=1}^n [|\mathbf{u}_s[l]|^2 \chi - \mathbf{z}_s[l] \log(|\mathbf{u}_s[l]|^2 \chi)],$$

where $\mathbf{u}_s = \mathbf{A}_s \mathbf{x}$.

The WF and TWF algorithms in [24] and [25] implement a direct minimization of $\mathcal{L}(\{\mathbf{u}_s\})$ based on straightforward calculations of the gradient of $\mathcal{L}(\{\mathbf{u}_s\}, \mathbf{x})$ with respect to $\mathbf{x} \in \mathbb{C}^n$.

Contrary to this approach we reformulate the problem as a constrained optimization $\min_{\mathbf{u}_s, \mathbf{x}} \mathcal{L}(\{\mathbf{u}_s\})$ subject to $\mathbf{u}_s = \mathbf{A}_s \mathbf{x}$, $s = 1, \dots, L$. The quadratic penalization of the constraints leads to the criterion

$$\begin{aligned} \mathcal{L}_1(\{\mathbf{u}_s\}, \mathbf{x}) &= \sum_{s=1}^L \sum_{l=1}^n [|\mathbf{u}_s[l]|^2 \chi - \mathbf{z}_s[l] \log(|\mathbf{u}_s[l]|^2 \chi)] + \\ &\frac{1}{\gamma_1} \sum_{s=1}^L \|\mathbf{u}_s - \mathbf{A}_s \mathbf{x}\|_2^2, \end{aligned} \quad (7)$$

where the weight parameter $\gamma_1 > 0$.

The iterative alternating minimization is used for optimization of (7) with respect to $\mathbf{u}_s \in \mathbb{C}^m$ and $\mathbf{x} \in \mathbb{C}^n$:

$$\{\hat{\mathbf{u}}_s^t\} = \arg \min_{\{\mathbf{u}_s\}} \mathcal{L}_1(\{\mathbf{u}_s\}, \hat{\mathbf{x}}^t), \quad (8)$$

$$\hat{\mathbf{x}}^{t+1} = \arg \min_{\mathbf{x}} \mathcal{L}_1(\{\hat{\mathbf{u}}_s^t\}, \mathbf{x}). \quad (9)$$

Note that $\mathcal{L}_1(\{\mathbf{u}_s\}, \mathbf{x})$ is additive with respect to $\mathbf{u}_s[l]$ and minimization on $\{\mathbf{u}_s\}$ is reduced to the scalar minimization on $\mathbf{u}_s[l]$ of the criterion

$$|\mathbf{u}_s[l]|^2 \chi - \mathbf{z}_s[l] \log(|\mathbf{u}_s[l]|^2 \chi) + \frac{1}{\gamma_1} |\mathbf{u}_s[l] - \mathbf{v}_s[l]|^2, \quad \mathbf{v}_s = \mathbf{A}_s \mathbf{x}. \quad (10)$$

It can be verified that this minimization on the complex-valued $\mathbf{u}_s[l]$ gives $\text{angle}(\mathbf{u}_s[l]) = \text{angle}(\mathbf{v}_s[l])$. Inserting this solution in (10) we obtain the criterion depending only on the absolute values $|\mathbf{u}_s[l]|$:

$$|\mathbf{u}_s[l]|^2 \chi - \mathbf{z}_s[l] \log(|\mathbf{u}_s[l]|^2 \chi) + \frac{1}{\gamma_1} |\mathbf{u}_s[l] - \mathbf{v}_s[l]|^2. \quad (11)$$

After differentiation on $|\mathbf{u}_s[l]|$ we obtain the quadratic equation for the optimal absolute values. A unique non-negative root of this quadratic equation is

$$\begin{aligned} \mathbf{b}_s[l] &= \\ &\frac{|\mathbf{v}_s[l]| + \sqrt{|\mathbf{v}_s[l]|^2 + 4\mathbf{z}_s[l](\gamma_1 \chi)^{-2}(1 + 1/(\gamma_1 \chi))/\chi}}{2(1 + 1/(\gamma_1 \chi))(\gamma_1 \chi)}. \end{aligned} \quad (12)$$

Then, the solution for (8) is of the form [35]:

$$\mathbf{u}_s[l] = \mathbf{b}_s[l] \exp(j \cdot \text{angle}(\mathbf{v}_s[l])), \quad (13)$$

In this solution the amplitude $\mathbf{b}_s[l]$ depends on both the observation \mathbf{z}_s and the amplitude of $\mathbf{v}_s[l]$.

Note, that for large $\gamma_1\chi \rightarrow \infty$ (noiseless case)

$$\mathbf{u}_s[l] \rightarrow \sqrt{\mathbf{z}_s[l]/\chi} \exp(j \cdot \text{angle}(\mathbf{v}_s[l]), s = 1, \dots, L. \quad (14)$$

Optimization of $\mathcal{L}_1(\{\mathbf{u}_s^t\}, \mathbf{x})$ with respect to $\mathbf{x} \in \mathbb{C}^n$ (for the problem (9)) leads to the minimum condition of the form $\partial\mathcal{L}_1(\{\mathbf{u}_s^t\}, \mathbf{x})/\partial\mathbf{x}^* = 0$ and to the normal least-squares equation for \mathbf{x}

$$\sum_{s=1}^L \mathbf{A}_s^H \mathbf{A}_s \mathbf{x} = \sum_{s=1}^L \mathbf{A}_s^H \mathbf{u}_s \quad (15)$$

and to the solution

$$\mathbf{x} = \left(\sum_{s=1}^L \mathbf{A}_s^H \mathbf{A}_s \right)^{-1} \sum_{s=1}^L \mathbf{A}_s^H \mathbf{u}_s. \quad (16)$$

For the Fraunhofer approximation of the forward wavefront propagation diffraction operator $\mathbf{A}_s = \mathbf{F}_s$, where \mathbf{F}_s is the discrete Fourier transform, and $\mathbf{A}_s^H \mathbf{A}_s = \mathbf{I}_{n \times n}$ provided $n = m$. Then (16) takes the form

$$\mathbf{x} = \frac{1}{L} \sum_{s=1}^L \mathbf{A}_s^H \mathbf{u}_s. \quad (17)$$

In general, the situation can be much more complex, in particular, because $\mathbf{A}_s^H \mathbf{A}_s$ are ill-conditioned due to the fact that the operators \mathbf{A}_s are a low-path filters suppressing high frequency components of the object \mathbf{x} (e.g. [4]) and in many cases $m < n$.

Then the solution of (15) can be found using iterations standard for a normal linear equation:

$$\mathbf{e}^{k+1} = \mathbf{e}^k - \beta \left(\sum_{s=1}^L \mathbf{A}_s^H \mathbf{A}_s \mathbf{e}^k - \sum_{s=1}^L \mathbf{A}_s^H \mathbf{u}_s \right), k = 0, 1, \dots, \quad (18)$$

where \mathbf{e}^k is an estimate of \mathbf{x} and $\beta > 0$ is a step-size parameter.

Note, that in this modeling $\mathbf{A}_s \mathbf{e}^k$ is the forward propagation of the wavefront \mathbf{e}^k and $\mathbf{A}_s^H \mathbf{A}_s \mathbf{e}^k$ is the backward propagation of the wavefront $\mathbf{A}_s \mathbf{e}^k$. The phase diversity required for phase retrieval formally means that $\sum_{s=1}^L \mathbf{A}_s^H \mathbf{A}_s$ is a positive definite matrix. Then, Eq.(15) has a unique solution (16) and exists a small enough β such that the iterations in (18) converge.

Combining the solutions (14) and (16) for Eqs.(8)-(9) we arrive to the iterative GS algorithm shown in Table 1.

At Step 1 the object wavefront estimate $\hat{\mathbf{x}}^t$ propagates using the operators \mathbf{A}_s and defines the wavefront $\hat{\mathbf{v}}_s^t$ at the sensor plane. At Step 2 this wavefront is updated to the variable $\hat{\mathbf{u}}_s^t$ by changing the amplitude of $\hat{\mathbf{v}}_s^t$ according

to the given observations \mathbf{z}_s , while the phase of $\hat{\mathbf{v}}_s^t$ is preserved in $\hat{\mathbf{u}}_s^t$. At Step 3 the estimates $\{\hat{\mathbf{u}}_s^t\}$ backpropagate to the object plane and update the object wavefront $\hat{\mathbf{x}}^{t+1}$.

We have here both typical features of the GS algorithms. First, iterative forward and backward propagations, and second, update of the amplitudes in the transform domain accordingly to the given observations. Respectively, we use the name GS for this phase retrieval algorithm.

Contrary to the usual heuristic design of this type of the algorithms this one is derived from optimization formulation. Remind that the assumption $\gamma_1\chi \rightarrow \infty$ is used in this design. It means that this algorithm is optimal for noiseless observations only. For noisy observations (χ is not large) the algorithm can be optimized accordingly using in Step 2 the amplitudes defined by (12) instead of the standard GS rule (14).

TABLE I
GS PHASE RETRIEVAL ALGORITHM

	Input: $\{\mathbf{z}_s\}, s = 1, \dots, L, \mathbf{x}^1;$
	For $t = 1, \dots, N;$
1.	Forward propagation:
	$\hat{\mathbf{v}}_s^t = \mathbf{A}_s \hat{\mathbf{x}}^t, s = 1, \dots, L;$
2.	Observation constrains:
	$\hat{\mathbf{u}}_s^t[l] = \sqrt{\mathbf{z}_s[l]/\chi} \exp(j \cdot \text{angle}(\hat{\mathbf{v}}_s^t[l])), s = 1, \dots, L;$
3.	Backward propagation:
	$\hat{\mathbf{x}}^{t+1} = (\sum_{s=1}^L \mathbf{A}_s^H \mathbf{A}_s)^{-1} \sum_{s=1}^L \mathbf{A}_s^H \hat{\mathbf{u}}_s^t;$
	Output: $\hat{\mathbf{x}}^{N+1}.$

With $\gamma_1 \rightarrow \infty$ the criterion $\mathcal{L}_1(\{\mathbf{u}_s\}, \mathbf{x}) \rightarrow \sum_{s=1}^L \sum_{l=1}^n [|\mathbf{u}_s[l]|^2 \chi - \mathbf{z}_s[l] \log(|\mathbf{u}_s[l]|^2 \chi)]$, thus the GS algorithm minimizes the minus loglikelihood written for Poissonian observations. Note that the TWF algorithm in [25] is designed for minimization of this latter criterion. Thus, both TWF and GS algorithms being completely different are intended for minimization of the same criterion.

It seems that the goal of the optimal algorithm design is achieved because the solution if found is optimal for Poissonian observations at least for large L . However, a large L is not practical and for a small L randomness of observations is revealed in randomness of the object wavefront \mathbf{x} , in particular in the object phase. The random

errors in \mathbf{x} can be quite strong due to the fact that the problem at hand is usually ill-posed and noise amplification for the variables of interest is a typical effect.

This point motivates us to move further and to developed an algorithm with improved filtering properties for \mathbf{x} . In what follows we show that the sparsity hypotheses for amplitude and phase of \mathbf{x} is a relevant instrument to achieve this goal.

B. Sparse phase retrieval

Two principally different variational formulations classified as the analysis and synthesis approaches can be viewed for sparse modelling. In the synthesis approach, the relations between the signal and spectrum variables are given by the synthesis equations (4), while in the analysis approach these relations are given by the analysis equations (5).

In the synthesis approach the variational setup is of the form

$$\begin{aligned} \min_{\boldsymbol{\theta}_a, \boldsymbol{\theta}_\varphi} \sum_{s=1}^L \sum_{l=1}^n [|\mathbf{u}_s[l]|^2 \chi - \mathbf{z}_s[l] \log(|\mathbf{u}_s[l]|^2 \chi)] + \\ \alpha_1 \|\boldsymbol{\theta}_a\|_0 + \alpha_2 \|\boldsymbol{\theta}_\varphi\|_0, \end{aligned} \quad (19)$$

$$\text{s.t. } \mathbf{u}_s = \mathbf{A}_s \mathbf{x}, \mathbf{x} = (\Psi_a \boldsymbol{\theta}_a) \circ \exp(j \Psi_\varphi \boldsymbol{\theta}_\varphi). \quad (20)$$

The first summand in (19) is the minus loglikelihood corresponding to the Poissonian distribution (6), where $\mathbf{y}_s[l]$ are replaced by $|\mathbf{u}_s[l]|^2$. The pseudo l_0 -norms with the coefficients α_1 and α_2 are included in order to enable the sparsity of the amplitude and phase in the spectrum domain. According to the constrains (20) $\mathbf{u}_s = \mathbf{A}_s [(\Psi_a \boldsymbol{\theta}_a) \circ \exp(j \Psi_\varphi \boldsymbol{\theta}_\varphi)]$. Substituting this expression in (19) one can see that the minimization criterion depends only the spectrum variables $\boldsymbol{\theta}_a$ and $\boldsymbol{\theta}_\varphi$. This is a specific feature of the synthesis formulation: minimization is produced with respect to the spectrum variables and only the synthesis operators Ψ_a and Ψ_φ are used. The amplitude and the phase are calculated due to Eqs.(4). Thus, \mathbf{x} is calculated using only the synthesis operators.

In the analysis approach the variational setup is of the form

$$\begin{aligned} \min_{\mathbf{x}} \sum_{s=1}^L \sum_{l=1}^n [|\mathbf{u}_s[l]|^2 \chi - \mathbf{z}_s[l] \log(|\mathbf{u}_s[l]|^2 \chi)] + \\ \alpha_1 \|\Phi_a \mathbf{b}\|_0 + \alpha_2 \|\Phi_\varphi \boldsymbol{\varphi}_{abs}\|_0, \end{aligned} \quad (21)$$

$$\text{s.t. } \mathbf{u}_s = \mathbf{A}_s \mathbf{x}, \mathbf{b} = \text{abs}(\mathbf{x}), \boldsymbol{\varphi}_{abs} = \mathcal{W}^{-1}(\text{angle}(\mathbf{x})). \quad (22)$$

Substituting \mathbf{u}_s , \mathbf{b} and φ_{abs} from (22) into (21) one can see that this criterion depends only on the spatial complex-valued variable \mathbf{x} . Thus, contrary to the synthesis approach the optimization is produced in the signal domain and only the analysis operators Φ_a and Φ_φ are used.

It is clear from (19)-(22) that both the synthesis and analysis setups lead to quite complex optimization problems.

Herein, we adopt a different Nash equilibrium approach. The constrained optimization with a single criterion function, as it is in (19) and (21), is replaced by a search for the Nash equilibrium balancing two criteria. Details of this approach, links with the game theory and demonstrations of its efficiency for the synthesis-analysis sparse inverse imaging can be seen in [29], where it is done for linear real-valued observation modeling. Applications of the Nash equilibrium technique for optical problems with nonlinear objects and nonlinear observations can be seen in [26], [27], [35]. In what follows we use the approaches developed in these papers.

C. SPAR algorithm

The following two criteria are introduced for formalization of the algorithm design:

$$\begin{aligned} \mathcal{L}_1(\{\mathbf{u}_s\}, \mathbf{x}) &= \sum_{s=1}^L \sum_{l=1}^n [|\mathbf{u}_s[l]|^2 \chi - \mathbf{z}_s[l] \log(|\mathbf{u}_s[l]|^2 \chi)] + \\ &\frac{1}{\gamma_1} \sum_{s=1}^L \|\mathbf{u}_s - \mathbf{A}_s \mathbf{x}\|_2^2, \end{aligned} \quad (23)$$

$$\begin{aligned} \mathcal{L}_2(\boldsymbol{\theta}_\varphi, \boldsymbol{\theta}_a, \varphi_{abs}, \mathbf{b}) &= \tau_a \cdot \|\boldsymbol{\theta}_a\|_0 + \tau_\varphi \cdot \|\boldsymbol{\theta}_\varphi\|_0 + \\ &\frac{1}{2} \|\boldsymbol{\theta}_a - \Phi_a \mathbf{b}\|_2^2 + \frac{1}{2} \|\boldsymbol{\theta}_\varphi - \Phi_\varphi \varphi_{abs}\|_2^2, \end{aligned} \quad (24)$$

where $\mathbf{b} = abs(\mathbf{x})$, $\varphi_{abs} = \mathcal{W}^{-1}(angle(\mathbf{x}))$.

The criterion (23) is identical to (7) and already discussed in Subsection III-A. As it is emphasized in Subsection II-A we use the separate sparse modeling for the absolute phase φ_{abs} and the amplitude \mathbf{b} of the wavefront \mathbf{x} . The criterion (24) promotes this sparsity in the transform domain. The regularization terms $\frac{1}{2} \|\boldsymbol{\theta}_a - \Phi_a \mathbf{b}\|_2^2$ and $\frac{1}{2} \|\boldsymbol{\theta}_\varphi - \Phi_\varphi \varphi_{abs}\|_2^2$ are squared Euclidean norms calculated for differences between spectra $\boldsymbol{\theta}_a$ and $\boldsymbol{\theta}_\varphi$ and their predictors $\Phi_a \mathbf{b}$ and $\Phi_\varphi \varphi_{abs}$.

Minimization of $\mathcal{L}_1(\{\mathbf{u}_s\}, \mathbf{x})$ with respect to \mathbf{u}_s and \mathbf{x} gives the solutions shown in (12) and (16).

Minimization of \mathcal{L}_2 on θ_a and θ_φ results in the well known hard-thresholding solutions:

$$\hat{\theta}_a = (\Phi_a \mathbf{b}) \circ 1 [abs(\Phi_a \mathbf{b}) \geq \sqrt{2\tau_a}], \quad (25)$$

$$\hat{\theta}_\varphi = (\Phi_\varphi \varphi_{abs}) \circ 1 [abs(\Phi_\varphi \varphi_{abs}) \geq \sqrt{2\tau_\varphi}],$$

where $1[\mathbf{w}]$, $\mathbf{w} \in \mathbb{R}^p$, is an element-wise vector function: $\mathbb{R}^p \mapsto \mathbb{R}^p$, $1[\mathbf{w}_k] = 1$ if $\mathbf{w}_k \geq 0$ and $1[\mathbf{w}_k] = 0$ if $\mathbf{w}_k < 0$.

Here $th_a = \sqrt{2\tau_a}$ and $th_\varphi = \sqrt{2\tau_\varphi}$ are thresholds for the amplitude and the phase, respectively. The items of the spectral coefficients $abs(\Phi_a \mathbf{b})$ and $abs(\Phi_\varphi \varphi_{abs})$, which are smaller than the corresponding thresholds are zeroed in (25).

It is useful to mention that in general multidimensional minimization of the l_0 -pseudonorm results in a hard computational problem. The considered criterion $\mathcal{L}_2(\theta_\varphi, \theta_a, \varphi_{abs}, \mathbf{b})$ is additive with respect to the items of the vectors θ_a , θ_φ . Then the minimization can be produced independently for each items of these vectors and this scalar optimization allows a simple analytical solution shown in (25).

According to the idea of the Nash equilibrium balancing multiple penalty functions (e.g. [36]) the proposed algorithm is composed of alternating optimization steps performed for the criteria \mathcal{L}_1 and \mathcal{L}_2 . It leads to the iterative procedure:

$$\{\hat{\mathbf{u}}_s^t\} = \arg \min_{\{\mathbf{u}_s\}} \mathcal{L}_1(\{\mathbf{u}_s\}, \hat{\mathbf{x}}^t), \quad (26)$$

$$\hat{\mathbf{x}}^t = \arg \min_{\mathbf{x}} \mathcal{L}_1(\{\hat{\mathbf{u}}_s^t\}, \mathbf{x}), \quad (27)$$

$$(\hat{\theta}_\varphi^t, \hat{\theta}_a^t) = \arg \min_{\theta_\varphi, \theta_a} \mathcal{L}_2(\theta_\varphi, \theta_a, \hat{\varphi}_{abs}^t, \hat{\mathbf{b}}^t), \quad (28)$$

$$\hat{\mathbf{b}}^{t+1} = \Psi_a \hat{\theta}_a^t, \hat{\varphi}_{abs}^{t+1} = \Psi_\varphi \hat{\theta}_\varphi^t \quad (29)$$

$$\hat{\mathbf{x}}^{t+1} = \hat{\mathbf{b}}^{t+1} \circ \exp(j\hat{\varphi}_{abs}^{t+1}), \quad (30)$$

where \mathbf{b}^{t+1} and φ_{abs}^{t+1} are updates of the amplitude and the absolute phase for \mathbf{x} .

The success of any sparse imaging depends on how reach and redundant are the transforms/dictionaries used for analysis and synthesis. In our algorithm for the analysis and synthesis operations we use the BM3D frames, where BM3D is the abbreviation for Block-Matching and 3D filtering [29].

Let us recall some basic ideas of this advanced technique. At the first stage the image is partitioned into small overlapping square patches and further the vector corresponding to each patch is modeled as a sparse linear

combination of vectors taken from a given orthonormal bases. For each patch a group of similar patches is collected which are stacked together and form a 3D array (group). This stage is called *grouping*. The entire 3D group-array is projected onto a 3D transform basis. The obtained spectral coefficients are hard-thresholded and the inverse 3D transform gives the filtered patches, which are returned to the original position of these patches in the image. This stage is called *collaborative* filtering. This process is repeated for all pixels of the entire wavefront and obtained overlapped filtered patches are aggregated in the final image estimate. This last stage is called *aggregation*. The details of BM3D as an advanced image filter can be seen in [37].

It follows from [29], that the steps (28)-(29) including the *grouping* operations defining the analysis Φ and synthesis Ψ frames can be combined in a single algorithm. In what follows we use the notation BM3D for this algorithm. Note, that the standard BM3D algorithm as it is presented in the original paper [37] is composed from two successive steps: thresholding and Wiener filtering. In this paper BM3D corresponding to the procedures (28) and (29) consists of only the first thresholding (hard-thresholding) step.

The criterion \mathcal{L}_2 is separable on the variables θ_φ and θ_a . It follows that the corresponding solutions can be calculated in parallel for the amplitude and the phase. Using the BM3D algorithm for implementation of the steps (28)-(29) we obtain:

$$\begin{aligned}\hat{\mathbf{b}}^{t+1} &= BM3D_{ampl}(\hat{\mathbf{b}}^t, th_a), \\ \hat{\varphi}_{abs}^{t+1} &= BM3D_{phase}(\hat{\varphi}_{abs}^t, th_\varphi).\end{aligned}\tag{31}$$

In (31) we use BM3D with different subscripts because different parameters can be used in BM3D for amplitude and phase processing.

Combining the solutions obtain in Subsection III-A for (26)-(27) and for (31) into the steps (28)-(29) we arrive to the phase retrieval algorithm shown in Table 2.

The first three steps of this algorithm are identical to the corresponding three steps of the GS algorithm with the only difference that in Step 2 instead of the standard GS rule we use the more general procedure (12) for calculation of the amplitude taking into consideration the noisiness of observations.

The sparsification (filtering on the base of sparse approximation) is produced in Step 5. The unwrapping of the phase in Step 4 is necessary in order to use the sparsity hypothesis imposed on the absolute phase. In the GS

TABLE II
SPAR PHASE RETRIEVAL ALGORITHM

	Input: $\{\mathbf{z}_s\}, s = 1, \dots, L, \mathbf{x}^1$;
	For $t = 1, \dots, N$;
1.	Forward propagation:
	$\hat{\mathbf{v}}_s^t = \mathbf{A}_s \hat{\mathbf{x}}^t, s = 1, \dots, L$;
2.	Poissonian noise suppression:
	$\hat{\mathbf{u}}_s^t = \hat{\mathbf{b}}_s^t \circ \exp(j \cdot \text{angle}(\hat{\mathbf{v}}_s^t))$, Eq.(12) for $\hat{\mathbf{b}}_s^t$;
3.	Backward propagation:
	$\hat{\mathbf{x}}^t = (\sum_{s=1}^L \mathbf{A}_s^H \mathbf{A}_s)^{-1} \sum_{s=1}^L \mathbf{A}_s^H \hat{\mathbf{u}}_s^t$;
4.	Phase unwrapping:
	$\hat{\varphi}_{abs}^t = \mathcal{W}^{-1}(\text{angle}(\hat{\mathbf{x}}^t))$;
5.	Phase and amplitude filtering:
	$\hat{\varphi}_{abs}^{t+1} = BM3D_{phase}(\hat{\varphi}_{abs}^t, th_\varphi)$,
	$\hat{\mathbf{b}}^{t+1} = BM3D_{ampi}(\text{abs}(\hat{\mathbf{x}}^t), th_a)$;
6.	Object wavefront update:
	$\hat{\mathbf{x}}^{t+1} = \hat{\mathbf{b}}^{t+1} \circ \exp(j\hat{\varphi}_{abs}^{t+1})$;
	Output: $\hat{\varphi}_{abs}^{N+1}, \hat{\mathbf{b}}^{N+1}$.

algorithm the update of \mathbf{x} is produced in complex domain. Contrary to it the update of \mathbf{x} in Step 6 is produced through the updates of the amplitude and the absolute phase calculated in parallel.

The GS algorithm proposed in Subsection III-A can be treated as a special case of the SPAR algorithm. Indeed, if $th_\varphi, th_a \rightarrow 0$ the BM3D filters do not filter input signals. Then, we do not need the absolute phase, Step 4 as well as Step 5 can be dropped and the SPAR algorithm becomes identical to the GS algorithm within the difference in Step 2.

IV. NUMERICAL EXPERIMENTS

For our simulations we select the coded diffraction pattern scenario (2):

$$\mathbf{y}_s = |\mathbf{F}\mathbf{D}_s \mathbf{x}|^2, s = 1, \dots, L. \quad (32)$$

Following to the publications [24] and [25] the wavefront modulation is enabled by the random phases ϕ_k in \mathbf{D}_s with equal probabilities taking four values $[0, \pi/2, -\pi/2, \pi]$ and the number of experiments $L = 12$. To some extent, the choice for experiments of the model (32) with these four random phase values for phase modulation is caused by our intention to consider TWF as the main counterpart to our algorithms. The rigorous mathematical background of TWR developed for the Poissonian data confirms that this algorithm one of the best in the area [25]. The MATLAB codes of TWF provided by the authors for the model (32) make the comparative analysis simple for implementation. While the results in [25] are presented mainly for noiseless data or small level noise herein we are concentrated on noisy data and show that in this case the developed sparse modeling allows to achieve a dramatic improvement in the accuracy of phase and amplitude imaging. For the model (32) $\mathbf{A}_s = \mathbf{F}\mathbf{D}_s$ and

$$\sum_{s=1}^L \mathbf{A}_s^H \mathbf{A}_s = \sum_{s=1}^L \mathbf{D}_s^H \mathbf{F}^H \mathbf{F} \mathbf{D}_s = \sum_{s=1}^L \mathbf{D}_s^H \mathbf{D}_s = L \mathbf{I}_{n \times n}, \quad (33)$$

and Step 2 of our algorithm is simplified to the form

$$\hat{\mathbf{x}}^t = \frac{1}{L} \sum_{s=1}^L \mathbf{D}_s^H \mathbf{F}^H \hat{\mathbf{u}}_s^t. \quad (34)$$

All results presented in this section can be reproduced by running the publicly available MATLAB demo-codes ¹.

A. Processing without phase unwrapping

For experiments we use 256×256 MATLAB test-images: Lena and Cameraman. For the phase they are scaled to the interval $[0, \pi/2]$. In this case the phase unwrapping is not required and Step 4 is omitted in the SPAR algorithm.

The accuracy of the wavefront reconstruction is characterized by *RMSE* criteria calculated independently for amplitude and phase. In the phase retrieval problem the object phase image can be estimated within an invariant phase-shift only. Following [24] and [25] the estimated phase image is corrected by an invariant phase-shift φ_{shift} defined as

$$\varphi_{shift} = \arg\left(\min_{\varphi \in [0, 2\pi]} \|\exp(-j\varphi)\hat{\mathbf{x}} - \mathbf{x}\|_2^2\right), \quad (35)$$

here \mathbf{x} and $\hat{\mathbf{x}}$ are the true wavefront and its estimate, respectively. This correction of the phase is done only for calculation of the criteria and for result imaging and is not used in the algorithm iterations.

¹<http://www.cs.tut.fi/~lasip/DDT/index3.html>

In what follows we compare the results obtained by TWF (MATLAB codes available at <http://web.stanford.edu/~yxchen/TWF>) and ours GS and SPAR algorithms. The results are shown as functions of the Poissonian scale parameter χ , $2.5 \cdot 10^{-5} \leq \chi \leq 1$. The smallest χ results in the noisiest data. The corresponding Signal-to-Noise Ratio (SNR) is calculated as

$$SNR = 10 \log_{10}(\chi^2 \sum_{s=1}^L \|\mathbf{y}_s\|_2^2 / \sum_{s=1}^L \|\mathbf{y}_s \chi - \mathbf{z}_s\|_2^2) \text{ dB}. \quad (36)$$

As an interesting parameter of the observed data we calculate also the mean values of $\{\mathbf{z}_s\}$, $N_{photon} = \sum_{s=1}^L \sum_{k=1}^n \mathbf{z}_s(k) / Ln$, i.e. the mean number of photons per pixel of the sensor. Variations of SNR and N_{photon} naturally depend on amplitude and phase images. In our experiments these variations approximately take values from 1 to 60dB for SNR and from 1 photon to $6 \cdot 10^4$ photons per pixel for N_{photon} .

The achieved accuracy in $RMSE$ values is shown in Figs. 2 and 4, respectively, for the phase images of Lena and Cameraman with the amplitude equal to 1.

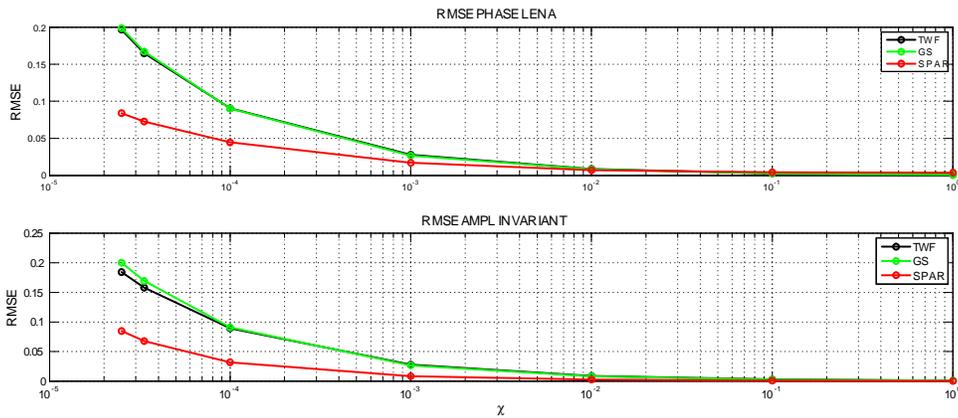


Fig. 2. Lena phase image: RMSE for phase and amplitude reconstructions versus the parameter χ . Comparison of the TWF, GS and SPAR algorithms.

Visualization of these phase reconstructions is presented in Figs.(3) and (5). The advantage of the SPAR algorithm is obvious both visually and numerically.

Now let us consider the more demanding scenario when both phase and amplitude are space varying. Lena is used for the phase and Cameraman for the amplitude. In these experiments Cameraman is scaled to the interval $[0.1, 1.1]$. The corresponding RMSE values are shown in Fig.6.



Fig. 3. Lena phase imaging for noisy data. Comparison of the TWF, GS and SPAR algorithms, $\chi = 2.5 \cdot 10^{-5}$, $SNR = 2.25 \text{ dB}$.

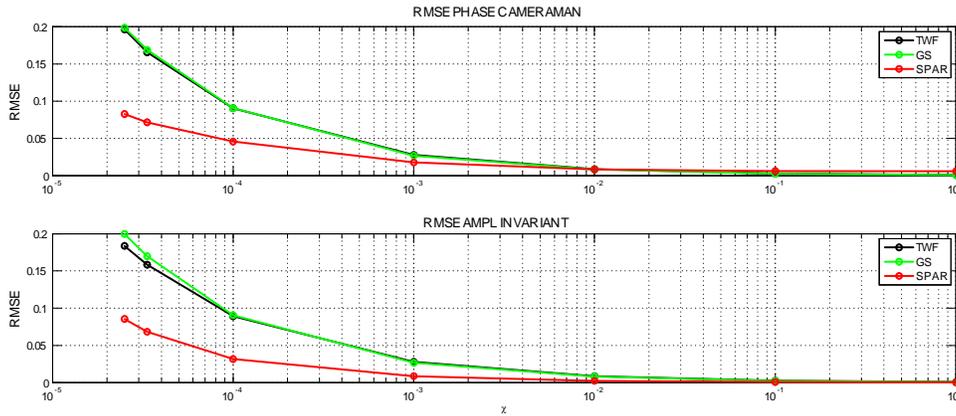


Fig. 4. Cameraman phase image: RMSE for phase and amplitude reconstructions versus the parameter χ . Comparison of the TWF, GS and SPAR algorithms.

It can be seen in comparison with the previous curves for RMSE that the accuracy of the reconstruction is significantly lower for the high-level noise, $\chi < 0.01$. Images in Fig.7 for noisy case, $\chi = 0.001$, show a visible degradation of the phase reconstruction due to the leakage from the space varying Cameraman amplitude. Some features of Cameraman can be seen in the Lena phase reconstruction. For the lower noise level, Fig.8, $\chi = 0.01$,



Fig. 5. Cameraman phase imaging for noisy data. Comparison of the TWF, GS and SPAR algorithms, $\chi = 2.5 \cdot 10^{-5}$, $SNR = 21 \text{ dB}$.

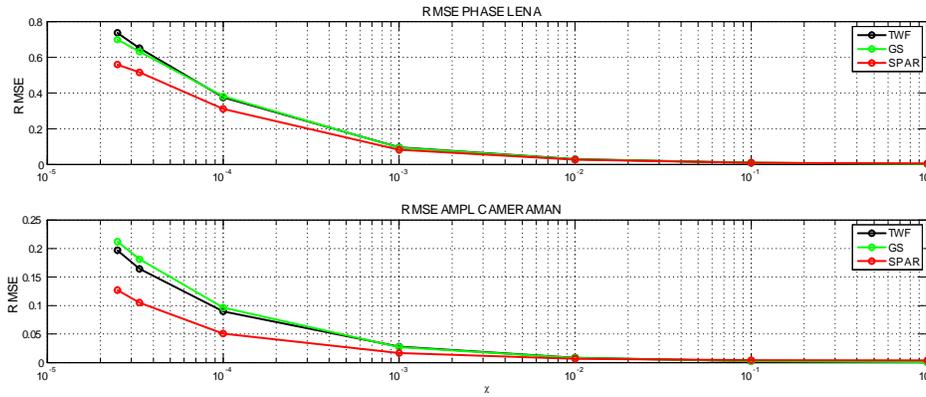


Fig. 6. Varying amplitude and phase modelled as Cameraman and Lena, respectively: RMSE for phase and amplitude reconstructions versus the parameter χ . Comparison of the TWF, GS and SPAR algorithms.

the reconstruction of the phase is of the quality close to the considered above cases when only phase is spatially varying and the amplitude is invariant. The serious advantage of the SPAR algorithm is obvious in these results.



Fig. 7. Varying amplitude and phase modelled as Cameraman and Lena, respectively: visualization of the phase reconstruction for the high level of the noise, $\chi = 0.001$. Traces of Cameraman are seen in these images.

B. Imaging from subsampled observations

Consider the case when only a part of the sensor measurements are used for image reconstruction. The problem is to reconstruct the high-resolution image from these subsampled data. It is a typical scenario in many optical applications where the sensor is able to register only a low frequency part of observations.



Fig. 8. Varying amplitude and phase modelled as Cameraman and Lena, respectively: visualization of the phase reconstruction for the comparatively low noise level, $\chi = 0.01$. Traces of Cameraman are not seen in these images.

In this case the criterion \mathcal{L}_1 in (23) takes the form

$$\begin{aligned} \mathcal{L}_1(\mathbf{u}_s, \mathbf{v}_s) &= \sum_{s=1}^L \sum_{l \in S} [|\mathbf{u}_s[l]|^2 \chi - \mathbf{z}_s[l] \log(|\mathbf{u}_s[l]|^2 \chi)] \\ &+ \frac{1}{\gamma_1} \sum_{s=1}^L \|\mathbf{u}_s - \mathbf{v}_s\|_2^2, \end{aligned} \quad (37)$$

where S denotes a set of the used sensor pixels.

Minimization of $\mathcal{L}_1(\mathbf{u}_s, \mathbf{v}_s)$ with respect \mathbf{u}_s results in Step 2 of the SPAR algorithm in the form

$$\hat{\mathbf{u}}_s[l] = \begin{cases} \mathbf{b}_s[l] \exp(j \cdot \text{angle}(\mathbf{v}_s[l])), & \text{if } l \in S, \\ \mathbf{v}_s[l] & \text{if } l \notin S. \end{cases} \quad (38)$$

Here, the vectors \mathbf{u} and \mathbf{v} have the same dimension, $\mathbf{u}, \mathbf{v} \in \mathbb{C}^n$, and only the pixels of \mathbf{u}_s from S are subject of filtering while others are equal to the corresponding elements of \mathbf{v}_s .

Similarly, Step 2 in the GS algorithm takes the form

$$\hat{\mathbf{u}}_s[l] = \begin{cases} \sqrt{\mathbf{z}_s[l]/\chi} \exp(j \cdot \text{angle}(\hat{\mathbf{v}}_s^t[l])) & \text{if } l \in S, \\ \mathbf{v}_s[l] & \text{if } l \notin S. \end{cases} \quad (39)$$

In the procedures (38)-(39) $\hat{\mathbf{u}}_s[l] = \mathbf{v}_s[l]$ for $l \notin S$ make a big difference with the Steps 2 in GS and SPAR algorithms.

Two types of subsampling are analyzed in our experiments. First, a random sampling of the sensor pixels and deterministic with a central square area of the sensor used for reconstruction. The last scenario means using only low frequency components of the Fourier transform of $\mathbf{D}_s \mathbf{x}$.

Denote as β a percentage of the pixels used for processing. Simulation experiments show that the modified GS and SPAR algorithms works very well provided subsampling upto $\beta = 25\%$, while the standard versions of these algorithms actually fail, in particular for the noisy data. Thus, the modifications proposed in (38)-(39) are of importance.

The dependence of RMSE on χ for the Lena phase modelling and invariant amplitude imaging are shown in Fig.9 for the deterministic subsampling with $\beta = 25\%$. Visualization of the phase imaging by SPAR and GS algorithms for $\varkappa = 0.01$ are shown Fig.10. The advantage of the SPAR algorithm versus the GS algorithm visually and numerically in RMSE values is obvious. These results confirms a serious advantage of the SPAR algorithm versus the GS algorithm for the downsampling scenario. The SPAR algorithm not only filter noise in the phase and amplitude imaging but also allows to reconstruct fine details of the images missed in subsampled observations. Simulation results for the random sampling are quite similar to presented in Figs.9-10 and are not shown here.

The TWF algorithm in the form presented in [25] is not applicable for the considered subsampling.

Note that the efficiency of the algorithm with subsampling means that the sensors of the size smaller than the image size can be exploited for phase retrieval.

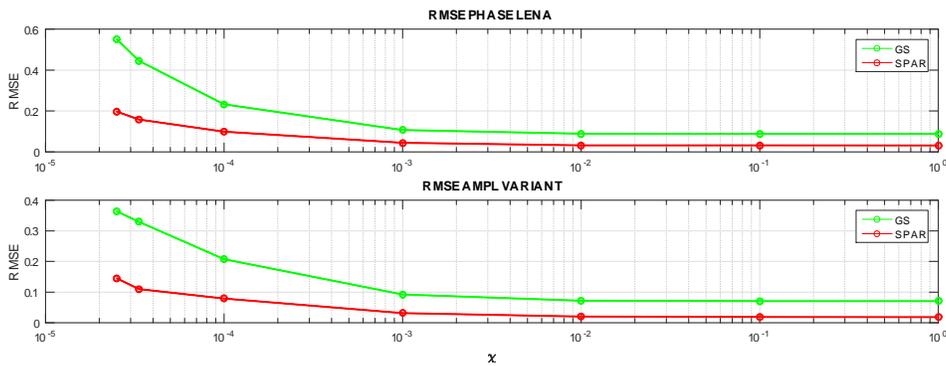


Fig. 9. Lena phase imaging for 25% subsampled observations: RMSE for phase and amplitude reconstructions versus the parameter χ . Comparison of the GS and SPAR algorithms.

C. Absolute phase imaging with phase unwrapping

Here we simulate three complex-valued data sets of the size 100×100 with the invariant amplitude equal to 1 and spatially varying absolute phase: Gaussian (phase range 44 radians), truncated Gaussian (phase range 44 radians), and Shear Plane (phase range 149 radians) exploited, in particular, in [32].



Fig. 10. Lena phase imaging: 25% subsampled noisy observations. Comparison of the SPAR and GS algorithms.

For the phase imaging we apply all three considered algorithms: TWF, GS and SPAR. The first two algorithms give the wrapped phase reconstructions which are unwrapped by the PUMA algorithm [38], identical to the one used in SPAR iterations.

The results are demonstrated for very noisy Poissonian observations obtained with $\varkappa = 2.5 \cdot 10^{-4}$. The 3D surfaces in Fig.11 are reconstructions of the absolute Gaussian phase. From left-to-right we can see three reconstructions TWF (failed), GS (quite noisy), SPAR and the true absolute phase surface. The SPAR reconstruction is slightly noisy and of the best quality and accuracy. In Fig.12 one can see the corresponding wrapped phases. RMSEs for the phase images in Figs.11-12 and in what follows are calculated for the absolute phase and wrapped phase, respectively.

Similar images in Figs.13-14 are shown for the truncated Gaussian surface. This discontinuous function is much more difficult for reconstruction and for unwrapping as compared with the continuous Gaussian surface. Again we can see that TWF fails, GS gives a low quality reconstruction while SPAR is able to show much better results.

Images in Figs.15-16 are shown for the discontinuous shear surface. We can see that TWF fails, GS gives a noisy reconstruction while the best results are demonstrated by the SPAR algorithm.

For lower noise level, $\chi > 10^{-3}$, all algorithms demonstrate a much better performance. TWF and GS show quite close results with a clear advantage of SPAR. For nearly noiseless cases, $\chi > 10^{-1}$, all algorithms enable a perfect reconstruction of all three absolute phase surfaces.

The advantage of SPAR versus GS reconstructions demonstrates the principal importance of the sparse phase

modeling in iterations of the SPAR algorithm. The unwrapping of the GS reconstruction applied only for the final estimate is not able to produce results comparable with the ones obtained by SPAR. For small noise level the accuracy of TWF and GS is nearly identical.

D. Parameters of the SPAR algorithm

The performance of the SPAR algorithm essentially depends on its parameters. Optimization can be produced for each magnitude/phase distribution and the noise level. However, in our experiments the parameters are fixed for all our experiments. The image patches in *BM3D* are square 8×8 . The group size is limited by 25 patches. The step size between the neighboring patches is equal to 3. The transforms DCT (for patches) and Haar (for the group length) are used for 3D group data processing in *BM3D*. In the shown results as an initial guess for the iterative GS and SPAR algorithm we use an image with the invariant amplitude equal to 1 and the zero mean Gaussian i.i.d. random phase with the standard deviation 0.1π .

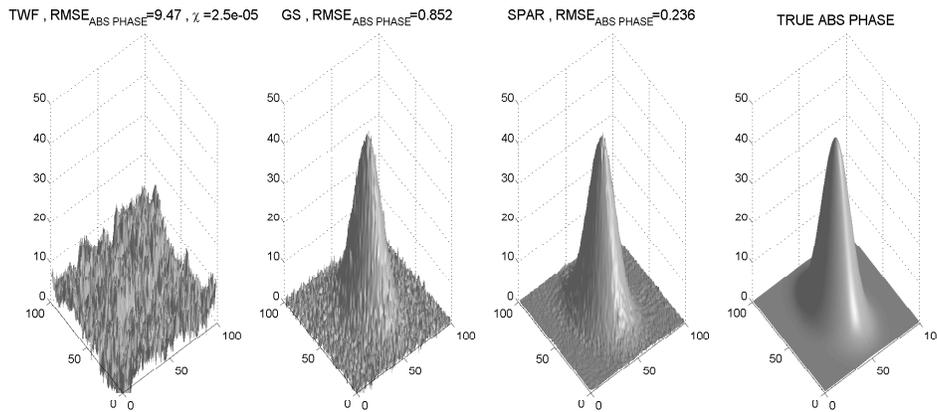


Fig. 11. The absolute (unwrapped) phase for the Gaussian phase object from noisy Poissonian observations. From left to right: TWF, GS, SPAR reconstructions and true phase.

The parameters defining the iterations of the algorithm are as follows: $\gamma_1 = 1/\chi$; $th_a = 1.4$; $th_\varphi = 1.4$. The number of the iterations is fixed to 50.

For our experiments we use MATLAB R2014a and the computer with the processor Intel(R) Core(TM) i7-4800MQ@ 2.7 GHz. The complexity of the algorithm is characterized by the time required for processing. For

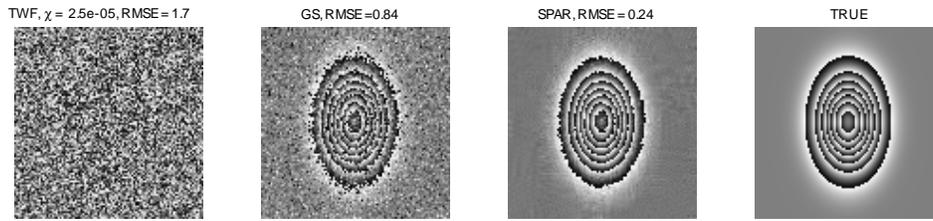


Fig. 12. The wrapped phase for the Gaussian phase object from noisy Poissonian observations. From left to right: TWF, GS, SPAR reconstructions and true wrapped phase.

50 iterations and 256×256 images this time is as follows: TWF \simeq 10 sec.; GS \simeq 5 sec.; SPAR \simeq 70 sec. (no unwrapping); SPAR \simeq 90 sec. (with unwrapping).

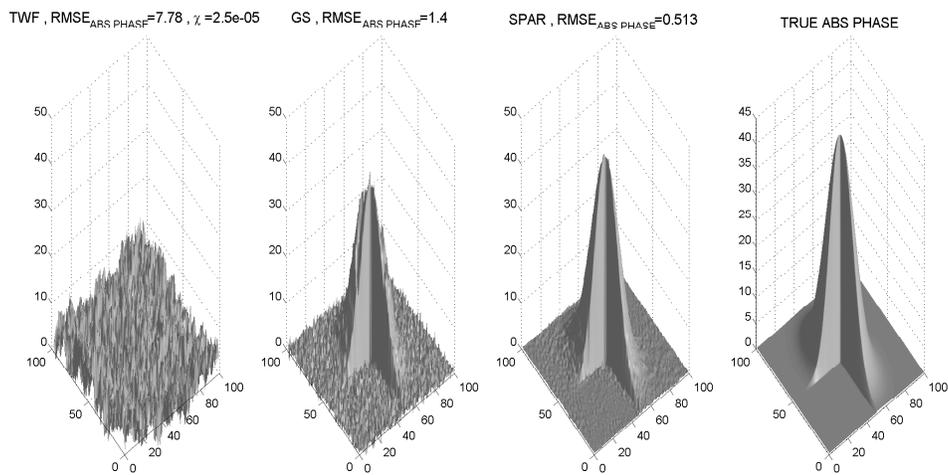


Fig. 13. The absolute (unwrapped) phase for the truncated Gaussian phase object from noisy Poissonian observations. From left to right: TWF, GS, SPAR reconstructions and true phase.

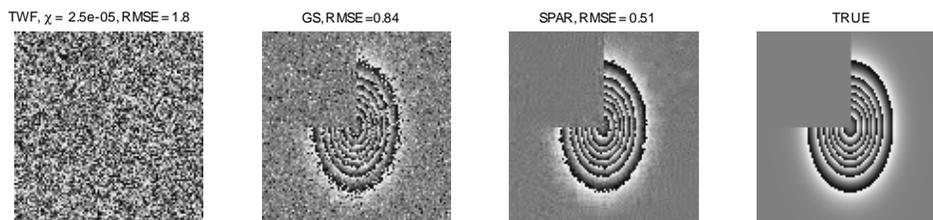


Fig. 14. The wrapped phase for the truncated Gaussian phase object from noisy Poissonian observations. From left to right: TWF, GS, SPAR reconstructions and true wrapped phase.

V. CONCLUSION

This paper introduces a variational approach to object phase and amplitude reconstruction from noisy Poissonian intensity observations in the typical phase retrieval scenario. The maximum likelihood criterion used in the developed multiobjective optimization (Nash equilibrium approach) defines the intention to reach statistically optimal estimates. The sparsity of amplitude and absolute phase is one of the key elements of the developed SPAR algorithm used for modeling spatially varying amplitude and phase. This sparse modeling enables regularization in general ill-posed inverse imaging problems. The complexity of the algorithm is defined by the built-in BM3D filters generating data adaptive synthesis and analysis frames varying in iterations. The efficiency of the algorithm is demonstrated by simulation experiments for the coded diffraction pattern scenario. The comparison is produced versus the state-of-the-art TWF algorithm. For noisy observations the SPAR algorithm demonstrates a definite advantage over TWF. For the low noise level the accuracy of the SPAR algorithm as well as its simplified version the GS algorithm is nearly identical to the accuracy of the TWF algorithm. The GS algorithm is essentially faster than TWF while SPAR computationally much more demanding is slower than both TWF and GS.

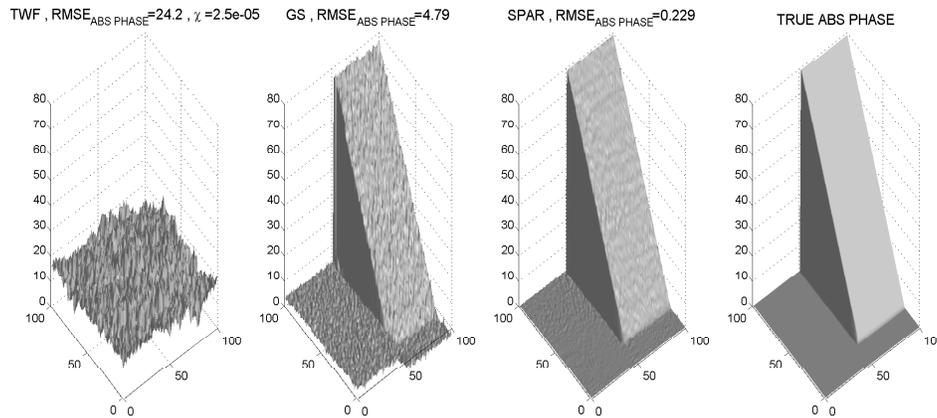


Fig. 15. The absolute (unwrapped) phase for the share plane phase object from noisy Poissonian observations. From left to right: TWF, GS, SPAR reconstructions and true phase.

REFERENCES

[1] D. L. Misell, “An examination of an iterative method for the solution of the phase problem in optics and electron optics: I. Test calculations,” *J. Phys. D Appl. Phys.*, vol. 6(18), 2200–2216 (1973).

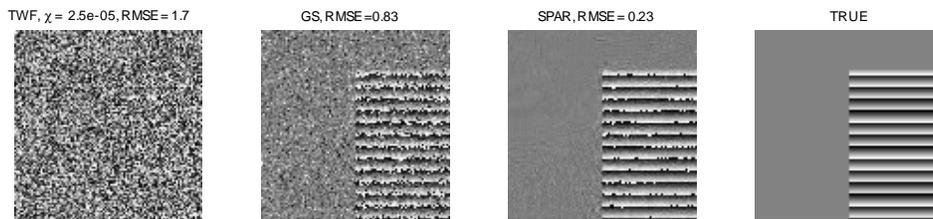


Fig. 16. The wrapped phase for the share plane phase object from noisy Poissonian observations. From left to right: TWF, GS, SPAR reconstructions and true wrapped phase.

- [2] W. O. Saxton, "Correction of artefacts in linear and nonlinear high resolution electron micrographs," *J. Microsc. Spectrosc. Electron.*, vol. 5, 661–670 (1980).
- [3] R. G. Paxman, T. J. Schulz, and J. R. Fienup, "Joint estimation of object and aberrations by using phase diversity," *J. Opt. Soc. Am. A*, vol. 9, no. 7, 1072–1085 (1992).
- [4] J. W. Goodman, *Introduction to Fourier optics*, 3rd Ed. (Roberts & Company, Englewood, 2005).
- [5] G. Pedrini, W. Osten, and Y. Zhang, "Wave-front reconstruction from a sequence of interferograms recorded at different planes," *Opt. Lett.*, vol. 30, 833–835 (2005).
- [6] P. Almero, G. Pedrini and W. Osten, "Complete wavefront reconstruction using sequential intensity measurements of a volume speckle field," *Appl. Opt.*, vol. 45, 8596–8605 (2006).
- [7] P. Almero, A. M. S. Maallo, and S. Hanson, "Fast-convergent algorithm for speckle-based phase retrieval and a design for dynamic wavefront sensing," *Appl. Opt.*, vol. 48, 1485–1493 (2009).
- [8] C. Kohler, F. Zhang, and W. Osten, "Characterization of a spatial light modulator and its application in phase retrieval," *Appl. Opt.*, vol. 48, no. 20, 4003–4008 (2009).
- [9] L. Camacho, V. Micy, Z. Zalevsky, and J. Garcha, "Quantitative phase microscopy using defocussing by means of a spatial light modulator," *Opt. Express.*, vol. 18, no.7, 6755–6766 (2010).
- [10] E. J. Candès, X. Lib, M. Soltanolkotabi, "Phase retrieval from coded diffraction patterns," *Appl. and Comp. Harm. Anal.*, vol.9, no. 2, pp. 277–299 (2015).
- [11] C. Falldorf, M. Agour, C. v. Kopylow, and R. B. Bergmann, "Phase retrieval by means of a spatial light modulator in the Fourier domain of an imaging system," *Appl. Opt.*, vol. 49, no. 10, 1826–1830 (2010).
- [12] V. Katkovnik and J. Astola, "Phase retrieval via spatial light modulator phase modulation in 4f optical setup: numerical inverse imaging with sparse regularization for phase and amplitude," *J. Opt. Soc. Am. A*, vol. 29, no 1, 105–116 (2012).
- [13] J. Glückstad, and D. Palima, "*Generalized phase contrast: applications in optics and photonics*," Springer Series in Optical Sciences, vol. 146 (2009).
- [14] R. W. Gerchberg and W. O. Saxton, A practical algorithm for the determination of phase from image and diffraction plane pictures, *Optik*, vol. 35, 237–246 (1972).

- [15] J. R. Fienup, Phase retrieval algorithms: a comparison, *Appl. Opt.*, vol. 21, 2758–2769 (1982).
- [16] C. Guo, S. Liu, Shi, J. T. Sheridan, "Iterative phase retrieval algorithms. I: optimization," *Appl. Opt.*, vol. 54, no.15, 4698-4708 (2015).
- [17] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev, "Phase retrieval with application to optical imaging: a contemporary overview," *IEEE Signal Processing Magazine*, 87-109 (2015).
- [18] E. J. Candes, Y. C. Eldar, T. Strohmer, and V. Voroninski, "Phase retrieval via matrix completion," *SIAM J. Imag. Sci.*, vol. 6, no. 1, 199–225 (2013).
- [19] Y. Shechtman, A. Beck, and Y. C. Eldar, "GESPAR: Efficient phase retrieval of sparse signals," *IEEE Trans. Image Process.*, vol. 62, no. 1–4, 928–938, (2014).
- [20] J.R. Fienup, "Phase-retrieval algorithms for a complicated optical system," *Appl. Opt.*, vol. 32, 1737-1746 (1993).
- [21] H. H. Bauschke, P. L. Combettes, and D. R. Luke, "Phase retrieval, error reduction algorithm, and Fienup variants: a view from convex optimization," *J. Opt. Soc. Am. A*, vol. 19, no. 7, 1334-1345 (2002).
- [22] R. Irwan and R.G. Lane, "Phase retrieval with prior information," *J. Opt. Soc. Am. A*, vol. 15, no. 9, pp. 2302-2311 (1998).
- [23] R.G. Lane, "Phase retrieval using conjugate gradient minimization," *J. Modern Optics*, vol. 8, 1797-1813 (1991).
- [24] E. J. Candès, X. Li and M. Soltanolkotabi, "Phase retrieval via Wirtinger flow: theory and algorithms," *IEEE Trans. on Information Theory*, vol. 61, no. 4, 1985–2007 (2015).
- [25] Y. Chen and E. J. Candès, "Solving random quadratic systems of equations is nearly as easy as solving linear systems," <http://statweb.stanford.edu/~candes/papers/TruncatedWF.pdf> (2015).
- [26] V. Katkovnik and J. Astola, "High-accuracy wavefield reconstruction: decoupled inverse imaging with sparse modeling of phase and amplitude," *J. Opt. Soc. Am. A*, vol. 29, 44 – 54 (2012).
- [27] V. Katkovnik and J. Astola, "Compressive sensing computational ghost imaging," *J. Opt. Soc. Am. A*, vol. 29, Is. 8, pp. 1556-1567 (2012).
- [28] V. Katkovnik and J. Astola, "Sparse ptychographical coherent diffractive imaging from noisy measurements," *J. Opt. Soc. Am. A*, vol. 30, 367-379 (2013).
- [29] A. Danielyan, V. Katkovnik, and K. Egiazarian, "BM3D frames and variational image deblurring," *IEEE Trans. Image Process.*, vol. 21, no 4, 1715 – 1728 (2012).
- [30] M. Elad, *Sparse and redundant representations: from theory to applications in signal and image processing*, Springer, 2010.
- [31] D. C. Ghiglia and M. D. Pritt, *Two-dimensional phase unwrapping: theory, algorithms, and software*. Wiley, 1998.
- [32] H. Hongxing, J. M. Bioucas-Dias, and V. Katkovnik, "Interferometric phase Image estimation via sparse coding in the complex domain," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 53, no. 5, 2587 - 2602 (2015).
- [33] V. Katkovnik, K Egiazarian, J. Bioucas-Dias, 'Phase imaging via sparse coding in the complex domain based on high-order SVD and nonlocal BM3D techniques,' *Proceedings of IEEE International Conference on Image Processing (ICIP 2014)*, 4587-4591 (2014).
- [34] Q. Kemao, "Two-dimensional windowed Fourier transform for fringe pattern analysis: Principles, applications and implementations," *Optics and Lasers in Engineering*, vol. 45, no. 2, 304 – 317 (2007).

- [35] V. Katkovnik, J. Bioucas-Dias, "Wavefront reconstruction in phase-shifting interferometry via sparse coding of amplitude and absolute phase," *J. Opt. Soc. Am. A*, vol. 31, no. 8, 1801-1810 (2014).
- [36] F. Facchinei and C. Kanzow, "Generalized Nash equilibrium problems," *4OR, Quart. J. Oper. Res.*, vol. 5, no. 3, 173–210 (2007).
- [37] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3D transform-domain collaborative filtering", *IEEE Trans. Image Process.*, vol. 16, no. 8, 2080-2095 (2007).
- [38] J. M. Bioucas-Dias and G. Valadão, " Phase unwrapping via graph cuts," *IEEE Trans. Image Process.*, vol. 16, no 3, 698–709 (2007).

Vladimir Katkovnik received Ph.D. and D.Sc. degrees in technical cybernetics from the Leningrad Polytechnic Institute (LPI), Leningrad, Russia, in 1964 and 1974, respectively. From 1964 to 1991, he was Associate Professor and then Professor with the Department of Mechanics and Control Processes, LPI. From 1991 to 1999, he was Professor with the Department of Statistics, University of South Africa, Pretoria, South Africa. From 2001 to 2003, he was Professor with the Kwangju Institute of Science and Technology, Gwangju, Korea. Since 2003, he is with the Department of Signal Processing, Tampere University of Technology, Tampere, Finland. He has published six books and over 300 papers. His research interests include stochastic signal processing, linear and nonlinear filtering, nonparametric estimation, computational imaging, nonstationary systems, and time-frequency analysis.