SOPR for sparse phase retrieval

Hieu Thao Nguyen∗† D. Russell Luke‡ Oleg Soloviev§ Michel Verhaegen¶

Abstract

We propose an algorithm, named SOPR, which is able to locally solve the sparse phase retrieval (SPR) problem given only one intensity point spread function (PSF) image. In abstract, the algorithm is a combination of the Gerchberg-Saxton method with the sparsity constraint but applied to the phase. The latter idea is newly developed in this paper. Convergence results for SOPR are rigorously established by showing that SOPR is equivalent to the cyclic projections method for solving an auxiliary possibly inconsistent feasibility problem involving three (locally) prox-regular sets and hence recent convergence theory for the cyclic projections method can be applied to SOPR. Extensions of SOPR for solving sparse phase retrieval given multiple intensity images can also be obtained naturally. Our numerical experiments for SPR show that our new regularization scheme which is implicitly incorporated in SOPR algorithm has clear advantages on solving SPR.

Keywords: sparse signal recovery, sparse phase retrieval, iterative algorithm, cyclic projections algorithm, nonconvex optimization, metric subregularity, prox-regularity, PSF intensity image.

1 Introduction

The phase retrieval is an inverse problem of recovering a complex signal from one or several measured intensity patterns. It appears in many scientific and engineering fields, including astronomy, crystallography, microscopy, optical manufacturing, and adaptive optics [10, 11, 24, 32]. Historically, there are three sorts of problems with three different names [4]. Phase retrieval refers to a basic form of image-based wavefront sensing in which a single image of a point source is used to estimate the unknown phase parameters of an optical system. Phase-diverse phase retrieval is similar in concept to phase retrieval except multiple images of a point source, each differing by a known phase term, are used to estimate the unknown phase parameters of an optical system. Phase diversity refers to a method of image-based wavefront sensing where multiple images of an unknown extended object or scene are used to estimate both the unknown phase parameters and the unknown object. The multiple images differ by some known phase terms. An important application of phase retrieval is to quantify the properties of an imaging system via its generalized pupil function (GPF). The fundamental advantage of this approach compared to those using intensity point spread functions (PSFs) or intensity optical transfer functions (OTFs) is that it is modifiable and automatically includes specific characterizations of the imaging system under investigation. The generic formulation of the phase retrieval problem is [32]:

\[
\text{find } x \in \mathbb{C}^n \text{ such that } |Mx|^2 = b + \varepsilon,
\]

where \( b \) is the measurement data, \( M \) is a known complex matrix of compatible sizes and \( \varepsilon \) is unknown measurement noise.

Depending on the parameterization, phase retrieval problem can be formulated in either the modal or the zonal form. In modal approach, one can either parameterize the phase aberration only [33], or
represent the GPF as a linear combination of some basis functions, e.g. the extended Nijboer-Zernike basis functions [6] and the radial basis functions, see for example [8,29]. In this formulation the dimension of the parameter variable $x$ does not depend on the size of the problem (the sizes of the sample grid) but rather the number of the basis functions in use to approximate the GPF. This makes the modal form approach a feasible choice for very large-scale problems in practice. The matrix $M$ in the modal form formulation in general has no desirable properties such as unitarity or isometry and hence convergence analysis of numerical methods for solving (1) is often challenging even in the noise-free case. It is an obvious prerequisite for the modal formulation that the GPF must be well approximated by a finite number of functions chosen from an orthogonal basis of an infinite dimensional space. This condition is not satisfied for example for the sparse phase retrieval considered in this paper. Complementing the modal formulation, the zonal form approach does not suffer the two potential drawbacks described above while requires the validity of the paraxial approximation in the imaging formulation. This formulation is theoretically based on the relationship that intensity PSF images are the squared magnitudes of the Fourier transforms of the GPF multiplied by well controlled phase diversity patterns [13]. The matrix $M$ in the zonal form formulation can hence be normalized to be unitary for problems given a single intensity image and isometric for those given multiple images. More importantly, available fast algorithms for evaluating the Fourier transform can be utilized to numerically compute solutions to phase retrieval problems in this Fourier transform-based model.

Dealing with sparse phase retrieval, we consider in this paper only the zonal form formulation. Since a PSF image captures the complex object GPF via the squared amplitude of its Fourier transform, multiple PSF images are, on the one hand, usually needed for retrieving the phase of the GPF up to a total piston term. On the other hand, simultaneously measuring several PSF images is not always a simple task for all applications in practice, especially for those involving real time imaging systems. A fundamental reason is that the use of double exposure procedures for such a measurement process would inevitably generate unwanted noise to the measured data. It is also known that under certain circumstances the phase can be estimated from a single intensity PSF image, see for example [7,12]. In this paper, we primarily focus on locally solving the sparse phase retrieval (SPR) given a single intensity PSF image, Section 3. Sparse phase retrieval with multiple intensity images is of course much easier to solve and will only be briefly discussed using the analysis at our disposal for the main research task.

For an overview of numerical methods for phase retrieval problems in general, we refer the reader to the papers [11,24,32]. For the purpose of this paper, we restrict ourselves to the discussion on algorithms which are applicable to the sparse phase retrieval problem (SPR) under investigation. By ignoring the information of sparsity, the Gerchberg-Saxton (GS) algorithm [12] technically fits the setting of SPR formulated in Section 3 and will be compared to our proposed algorithm in Section 7. To continue the discussion it is essential to figure out the key feature of SPR. The sparse phase retrieval problem is generally not a sparse signal recovery problem (SSR) given the amplitude of its Fourier transform. In the first problem, the sparsity is imposed on the phase of the signal whereas the signal itself may vanish nowhere (with respect to the canonical basis of $\mathbb{C}^{n \times n}$). Mathematically, a signal $\chi \mapsto \exp(j\Phi)$ can vanish nowhere no matter how sparse is its phase $\Phi$. Hence in this paper we make a distinction between the two problems SPR and SSR. A modification of the Fienup algorithm [10] for solving SSR was suggested in [26]. We demonstrate that to some extent this algorithm can be adapted for solving SPR. This algorithm will be compared to our algorithm along the discussion regarding solvability and sensitivity.

In the signal recovery literature, when the signal is real and sparse, the authors of [28,31] proposed and analyzed various iterative algorithms. Their rigorous treatment provides some mathematical hints for the study of SPR. However, it is emphasized again that there has been no concrete reformulation of SPR as an SSR problem. This fundamental fact is the barrier for adapting sparse signal retrieval algorithms in [28,31] for SPR and hence a comparison of our algorithm to these methods seems not feasible.

In this paper, we propose an iterative algorithm, named SOPR, for solving SPR and analyze its convergence. SOPR is technically a combination of the Gerchberg-Saxton method and the $\|\cdot\|_\infty$-regularization which is applied to the phase instead of the elements of the underlying space. This regularization scheme is newly developed in this paper. Convergence results for SOPR are rigorously established by showing that SOPR is equivalent to the cyclic projections algorithm for solving an auxiliary possibly inconsistent feasibility problem involving three (locally) prox-regular [30] sets and hence recent convergence theory developed in [25, Section 3.1] for the cyclic projections method can be applied to SOPR. Our numerical experiments for SPR comparing SOPR with the Gerchberg-Saxton method and the sparse Fienup algorithm [26] demonstrate that our new regularization scheme has clear advantages.
The remainder of paper is organized as follows. Basic mathematical notation will be introduced in Section 2. In Section 3 the sparse phase retrieval problem is formally formulated. In Section 4 we discuss the sparse feasibility approach for sparse signal retrieval and investigate a special setting of SPR where the sparse signal recovery algorithm [26] can be applied. In Section 5 we present the SOPR algorithm and discuss its significant features. In Section 6 convergence analysis for SOPR are rigorously established. Section 7 is devoted to demonstrate better numerical performance of SOPR compared to the relevant algorithms in particular and the advantages of our regularization scheme in general. Concluding remarks and discussion on future research are presented in Section 8.

2 Mathematical notation

The underlying space in this paper is a complex Hilbert space $\mathcal{H} = \mathbb{C}^{n \times n}$. The Frobenius norm is denoted $\| \cdot \|_F$. The 0-norm $\| \cdot \|_0$ of an object is the number of its nonzero entries. The element-wise multiplication is denoted $\odot$. The element-wise division $\oslash$, the element-wise absolute value $| \cdot |$ and the element-wise square root $\sqrt{\cdot}$ operations are also frequently used in this paper but without need for extra notation. $\text{Re}(\cdot)$ and $\text{Im}(\cdot)$ denote the real and the imaginary parts respectively of a complex number. The distance to a set $\Omega \subset \mathcal{H}$ is defined by

$$\text{dist}(\cdot, \Omega) : \mathcal{H} \rightarrow \mathbb{R}_+: x \mapsto \inf_{w \in \Omega} \| x - w \|_F$$

and the set-valued mapping

$$P_\Omega : \mathcal{H} \rightrightarrows \Omega : x \mapsto \{ w \in \Omega \mid \| x - w \|_F = \text{dist}(x, \Omega) \}$$

is the corresponding projection operator. A selection $w \in P_\Omega(x)$ is called a projection of $x$ on $\Omega$. An iterative sequence $x_{k+1} \in T(x_k)$ generated by an operator/algorithm $T : \mathcal{H} \rightrightarrows \mathcal{H}$ is said to converge $R$-linearly to $x^\ast$ with rate $c \in [0, 1)$ if there is a constant $\gamma > 0$ such that

$$\|x_k - x^\ast\|_F \leq \gamma c^k \quad \forall k \in \mathbb{N}.$$  

3 Sparse phase retrieval

We consider the sparse phase retrieval problem (SPR) of

\begin{equation}
\text{finding} \quad x = \chi \odot \exp \left( j (\Phi^d + \Phi) \right) \in \mathbb{C}^{n \times n} \\
\text{such that} \quad |F(x)|^2 = b + \varepsilon, \tag{2}
\end{equation}

where $\chi \in \mathbb{R}_+^{n \times n}$ is a known amplitude distribution, $\Phi^d \in \mathbb{R}^{n \times n}$ is a known phase diversity, $\Phi \in \mathbb{R}^{n \times n}$ is the unknown variable which is known a priori to be sparse, $F$ is the (discrete) 2-dimensional Fourier transform, $b \in \mathbb{R}_+^{n \times n}$ is the measured intensity image of $x$, and $\varepsilon \in \mathbb{R}^{n \times n}$ is some unknown noise.

In this paper, $\chi$ is assumed to be a uniform distribution. For the simplicity of terminology, when $x = \chi \odot \exp \left( j (\Phi^d + \Phi) \right)$ is a solution to (2), $\Phi$ in the sequel will be called a phase solution to that problem. Note that the phase diversity $\Phi^d$ has an important role in applications though it brings no differences to (2) in terms of mathematics.

Since $\Phi$ is the only unknown of the above problem, the model of sparse phase retrieval given multiple intensity images is simply a combination of several models (2) each of which corresponds to a different phase diversity $\Phi^d$.

Despite of the sparsity of $\Phi$, problem (2) is not a sparse signal retrieval (SSR) [26, 28, 31] because the signal $x$ indeed vanishes nowhere. Hence, there is a profound difference between SPR and SSR. The link connecting the two problems established in Section 4 as well as the $\| \cdot \|_0$-regularization for SSR presented in Section 5 are novel to the best of our awareness.

4 $\| \cdot \|_0$-regularization for sparse signal retrieval

As discussed in Section 3, in general there is no concrete interpretation of SPR as a sparse signal retrieval (SSR) in the way that algorithms for the later problem can be applied to SPR. In this section
we establish a new link between SPR and SSR. As a by-product, we to some extent can compare our proposed algorithm to a sparse signal recovery method, see Section 7.

Let us consider the optimization problem with sparsity and complex-amplitude constraints (SOCA)

\[
\min_{x' \in \mathbb{C}^{n \times n}} \text{dist}^2(x', B) \quad \text{subject to} \quad \|x'\|_0 \leq \hat{s},
\]

where

\[
B := \{ x \in \mathbb{C}^{n \times n} \mid |\mathcal{F}(x)|^2 = b' \}
\]

for a given data \( b' \in \mathbb{R}^{n \times n} \), and \( \hat{s} \) is some sparsity parameter.

We note that SPR is not explicitly related to SOCA since the variable \( x \) in (2) is nonzero everywhere. However, the following observation following from the properties of the Fourier transform gives rise to a link between SPR and SOCA in the special setting.

**Lemma 4.1.** In the setting of SPR (2) with \( \Phi^d = 0 \), the equality

\[
|\mathcal{F}(x - \chi)| = |\mathcal{F}(x)|
\]

holds true for all entries except at the one corresponding to the zero frequency \( \xi_0 \). If in addition we suppose further that

\[
\text{Im}(\text{mean}(x)) = 0,
\]

then it also holds that

\[
|\mathcal{F}(x - \chi)(\xi_0)| = \|\chi\|_F - \mathcal{F}(x)(\xi_0).
\]

**Proof.** We have by the linearity of \( \mathcal{F} \) that

\[
\mathcal{F}(x - \chi) = \mathcal{F}(x) - \mathcal{F}(\chi).
\]

Since \( \chi \) is uniform, \( \mathcal{F}(\chi) \) is zero everywhere except at the zero frequency \( \xi_0 \). As a result, the first part of Lemma 4.1 follows. For the later part regarding the zero frequency, we first note that \( \mathcal{F}(x)(\xi_0) = \text{mean}(x) \), the average of the signal \( x \), and \( \mathcal{F}(\chi)(\xi_0) = \|\chi\|_F \) by the basic properties of the Fourier transform. Condition (6) then implies that \( \mathcal{F}(x)(\xi_0) = \text{Re}(\mathcal{F}(x)(\xi_0)) \). Since the (element-wise) argument of \( x \) is sparse and \( \chi \) is a uniform distribution, \( \text{Re}(\mathcal{F}(x)(\xi_0)) \) is positive. As a consequence of the Rayleigh-Parseval theorem we get \( |\mathcal{F}(x)(\xi_0)| \leq \|\chi\|_F \). Putting these facts together, we obtain

\[
|\mathcal{F}(x - \chi)(\xi_0)| = |\mathcal{F}(x)(\xi_0) - \|\chi\|_F| = \|\chi\|_F - \mathcal{F}(x)(\xi_0),
\]

which is (5) as claimed. \( \blacksquare \)

The implication of Lemma 4.1 is that under condition (6) solutions to SOCA (3) can be used as approximate solutions to SPR (2) as explained next.

**Lemma 4.2.** In the setting of SPR (2) with \( \Phi^d = 0 \), suppose condition (6) holds true and define \( b' \) by

\[
b'(\xi) = \begin{cases} 
\frac{b(\xi)}{\|\chi\|_F - \sqrt{b(\xi_0)}}^2 & \text{if } \xi \neq \xi_0, \\
\frac{b(\xi)}{2} & \text{if } \xi = \xi_0.
\end{cases}
\]

Let \( x = \chi \circ \exp(j\Phi) \in \mathbb{C}^{n \times n} \). Suppose that \( x' := x - \chi \) is a solution to SOCA (3) and \( |\mathcal{F}(x')|^2 = b' + \varepsilon' \) for a sufficiently small noise term \( \varepsilon' \). Then up to a noise term \( x \) is a solution to SPR.

**Proof.** By the assumption, \( x' \) is a solution to SOCA and it holds that

\[
|\mathcal{F}(x - \chi)|^2 = |\mathcal{F}(x')|^2 = b' + \varepsilon'.
\]

Then using Lemma 4.1, we have

\[
|\mathcal{F}(x)(\xi)|^2 = \begin{cases} 
|\mathcal{F}(x - \chi)(\xi)|^2 = b'(\xi) + \varepsilon'(\xi) & \text{if } \xi \neq \xi_0, \\
\|\chi\|_F - |\mathcal{F}(x - \chi)(\xi_0)|^2 & \text{if } \xi = \xi_0.
\end{cases}
\]

4
where the last equality follows from (7) and (8). Hence up to a noise term we can write $|F(x)|^2 = b + \varepsilon$. That is, $x$ is a solution to SPR and hence the proof is complete.

In the setting of Lemmas 4.1 and 4.2, one can approximately calculate the measurement $b' := |F(x')|^2$ of the sparse signal $x'$ from the given data $b = |F(x)|^2$ of SPR. This allows us to approach SPR using analysis techniques for SOCA. Note that the zero norm $\|x'\|_0$ in (3) is a regularization term encoding the sparsity nature of $x'$, equivalently $\Phi$, which is not captured in (2), and hence has no affects on the relationship between SPR and SOCA.

We consider SOCA as an extension of the sparse affine optimization problem and adapt the feasibility approach [15] for solving SOCA. To begin, let us assume to have a good guess $\hat{s}$ of the sparsity of solutions to (3) and consider the sparse feasibility problem (SFP) of

$$
\text{finding} \quad x' \in A \cap B,
$$

where $A := \{x \in \mathbb{C}^{n \times n} : \|x\|_0 \leq \hat{s}\}$ with the sparsity parameter $\hat{s}$, and the set $B$ is defined by (4). When the measurement data $b'$ is corrupted with noise, SFP may become an inconsistent feasibility problem and its solutions are accordingly understood as best approximation points in a reasonable sense for the two sets, see [23].

In the noise-free case and $\hat{s} = s$, the sparsity of the sparsest solutions to SOCA, the two problems SOCA and SFP are equivalent. In practice the sparsity $s$ is hardly known in advance. However, similar to the sparse affine feasibility [15], the sparsity parameter $\hat{s}$ is not numerically sensitive with respect to various fundamental feasibility algorithms. That is, for a relatively wide range of values of $\hat{s}$, these algorithms perform very much in the same nature (see also Section 7.2 for SOPR algorithm). Prominent advantages of the feasibility approach for SOCA are its low computational complexity compared to convex relaxation schemes for (3) and that convergence of solution methods can be understood and anticipated from the geometry of the feasibility problem.

The fundamental step of solving (9) by projection algorithms is to calculate the two projection operators $P_A$ and $P_B$. Let us define on $\mathbb{C}^{n \times n}$ the constraint set

$$
C := \{y \in \mathbb{C}^{n \times n} : |y|^2 = b'\}.
$$

Since $F$ is a unitary transformation, the projector $P_B$ takes the form

$$
P_B(x) = F^{-1}(P_C(F(x))) \quad \forall x \in \mathbb{C}^{n \times n},
$$

where $F^{-1}$ is the inverse 2-dimensional Fourier transform. As the set $C$ is geometrically the cartesian product of a number of circles of the complex plane, the explicit form of the projector $P_C$, which is in general a set-valued mapping, is available [2, 24]. From the viewpoint of numerical calculation, we only need a selection of the projector $P_C$. From now on, we only work with the following selection of $P_C$ which is, with a slight abuse of notation, also denoted by $P_C$.

$$
P_C(y) = \sqrt{b'} \odot \frac{y}{|y|} \quad \forall y \in \mathbb{C}^{n \times n}
$$

with use of the extended arithmetic operation $\frac{0}{0} = 1$. The corresponding selection of $P_B$ is

$$
P_B(x) = F^{-1}\left(\sqrt{b'} \odot \frac{F(x)}{|F(x)|} \right) \quad \forall x \in \mathbb{C}^{n \times n}.
$$

There is also a closed form for $P_A$. For each $x \in \mathbb{C}^{n \times n}$, let us denote $\mathcal{I}_{\hat{s}}(x)$ the set of all $\hat{s}$-tuples of indices of $\hat{s}$ largest in absolute value entries of $x$. The set $\mathcal{I}_{\hat{s}}(x)$ can contain multiple such $s$-tuples. The projector $P_A$ can be explicitly written [3] as follows: for any $x \in \mathbb{C}^{n \times n}$,

$$
P_A(x) = \{z \in \mathbb{C}^{n \times n} : \exists \mathcal{I} \in \mathcal{I}_{\hat{s}}(x) \quad \text{such that} \quad z \text{ equals } x \text{ on } \mathcal{I} \text{ and vanishes elsewhere}\}.
$$

In summary, the well known sparsity regularization for sparse signal retrieval described in this section can be applied to SPR in the special setting specified in Lemma 4.1. For the general setting of SPR (2), we propose a new strategy of $\|\cdot\|_0$-regularization which is implicitly incorporated in the SOPR algorithm described in the next section.
5 SOPR algorithm

This section contains our seemingly new idea for SPR. Under the assumption that the amplitude distribution \( \chi \) is uniform, a sparsity constraint can be applied on the phase to encode the sparsity property of the phase solution to SPR (2). The rigorous analysis of this technique will be provided in Section 6.

In the sequel, by saying “largest/smallest entries” of a vector or a matrix we always mean “largest/smallest in absolute value entries” to avoid repetition.

Algorithm 5.1. [SOPR algorithm for SPR problem (2)]

Input:
- \( b \in \mathbb{R}^{n \times n} \) — intensity PSF image
- \( \hat{s} \) — sparsity parameter
- \( \Phi_0 \) — initial guess for \( \Phi \)
- \( \tau \) — tolerance threshold.

Iteration process: given \( \Phi_k \)

1. \( x_k = \chi \circ \exp(j\Phi_k) \)
2. \( X_k = F(x_k \circ \exp(j\Phi^d)) \) — phase diversity and Fourier transform
3. \( Y_k = \sqrt{b} \circ \frac{X_k}{|X_k|} \) — amplitude constraint
4. \( y_k = |y_k| \circ \exp(j\Psi_k) = \exp(-j\Phi^d) \circ F^{-1}(Y_k) \) — inverse Fourier transform and phase correction
5. \( z_k = \chi \circ \exp(j\Psi_k) \) — amplitude constraint
6. find \( I_k \), a set of indices corresponding to \( \hat{s} \) largest entries of \( \Psi_k \)
7. \( \Phi_{k+1} = \Psi_k \) on \( I_k \) and zero elsewhere.

Stopping criterion: \( \|\Phi_k - \Phi_{k+1}\| < \tau \) or \( k \) exceeds a number of iterates.

Output: \( \hat{\Phi} = \Phi_{end} \) — the estimate.

Algorithm 5.1 is fundamentally a combination of the Gerchberg-Saxton (GS) algorithm (steps (1)–(5)) and a sparsity regularization but applied to the phase (steps (6)–(7)). When applied to SPR, this extra regularization step brings significant features of SOPR compared to the original GS method, see Section 7.

SOPR can be naturally adapted for solving SPR with multiple intensity images in the same way that the original GS method has been extended to various phase retrieval algorithms using phase diversity patterns. More specifically, steps (2)–(5) of SOPR can be enhanced by any algorithmic schemes for solving feasibility problems such as the alternating projections, the Douglas-Rachford method, the RAAR algorithm [22], the \( T_\lambda \) algorithm [27] and cyclic Douglas-Rachford scheme [5]. In order to demonstrate the effectiveness of our regularization strategy, in Section 7.5 we briefly compare the cyclic projections algorithm with the corresponding cyclic version of SOPR in terms of numerical performance for SPR with multiple intensity images.

6 Convergence analysis

In this section we will establish local convergence results for SOPR by showing its equivalence to the cyclic projections algorithm involving three sets. As a result, a comprehensive convergence theory for the cyclic projections method recently developed in [25] can be applied to SOPR. It is worth mentioning that the above mentioned theory also encompasses the inconsistent feasibility which in the setting of this paper corresponding to SOPR applied to SPR with noise. To avoid quite formidable technical details inherently arising for inconsistent feasibilities, we will present the convergence results in the noise-free setting.
To begin, let us define on $C^{n \times n}$ the three sets as follows:

$$
\Omega_1 := \left\{ x \in C^{n \times n} \mid |F(x \circ \exp(j\Phi^d))| = \sqrt{b} \right\}, \\
\Omega_2 := \left\{ x \in C^{n \times n} \mid |x| = \chi \right\}, \\
\Omega_3 := \left\{ x \in \Omega_2 \mid \|\arg(x)\|_0 \leq \delta \right\},
$$

where $\arg(x)$ denotes the element-wise argument of $x$ taking values in $(-\pi, \pi]$.

The next lemma shows that SOPR is nothing else but the cyclic projections method $P_{\Omega_1} P_{\Omega_2} P_{\Omega_3}$.

**Lemma 6.1.** The following statements regarding Algorithm 5.1 hold true.

(i) The combination of steps (2), (3) and (4) of SOPR amounts to $y_k \in P_{\Omega_1}(x_k)$.

(ii) Step (5) of SOPR amounts to $z_k \in P_{\Omega_2}(y_k)$.

(iii) The combination of steps (6), (7) and (1) of SOPR amounts to $x_{k+1} \in P_{\Omega_3}(z_k)$.

As a consequence,

$$
x_{k+1} \in \text{SOPR}(x_k) = P_{\Omega_1} P_{\Omega_2} P_{\Omega_3}(x_k) \quad \forall k \in \mathbb{N}.
$$

**Proof.** (i) We first observe that

$$
\Omega_1 = \{ x \in C^{n \times n} \mid x \circ \exp(j\Phi^d) \in C \},
$$

where the constraint set $C$ is given by (10).

Let us denote the linear operator $D : C^{n \times n} \to C^{n \times n}$ by

$$
x \mapsto D(x) := x \circ \exp(j\Phi^d) \quad \forall x \in C^{n \times n}.
$$

It is clear that $D$ is a unitary transform since the energy of any signal in $C^{n \times n}$ is invariant with respect to $D$. Its inverse clearly takes the form $D^{-1}(x) = x \circ \exp(-j\Phi^d)$. Now using (12) and the formula (11) with the unitary transform $F \circ D$ in place of $F$, we obtain

$$
P_{\Omega_1}(x) = D^{-1} \circ F^{-1} (P_C (F \circ D(x))) \quad \forall x \in C^{n \times n}.
$$

Comparing to steps (2), (3) and (4) of SOPR and using (14), we have that

$$
y_k = \exp(-j\Phi^d) \circ F^{-1} \left( \sqrt{b} \circ \frac{X_k}{|X_k|} \right) = \exp(-j\Phi^d) \circ F^{-1} \left( \sqrt{b} \circ \frac{F(x_k \circ \exp(j\Phi^d))}{|F(x_k \circ \exp(j\Phi^d))|} \right) = D^{-1} \circ F^{-1} (P_C (F \circ D(x_k))) = P_{\Omega_1}(x_k) \quad \forall k \in \mathbb{N}.
$$

The proof is complete.

(ii) This part is clear from step (5) of SOPR and the definition of $\Omega_2$.

(iii) This statement is rather unexpected since the $\| : \|_0$-regularization applied to the phase object on $(-\pi, \pi)^n$ turns out to be equivalent to a metric projection on the underlying space $C^{n \times n}$. Let $\Psi_k$, $I_k$ and $\Phi_{k+1}$ be as in steps (6), (7) and (1) of SOPR. We need to show that $x_{k+1} = \chi \circ \exp(j\Phi_{k+1})$ is a projection of $z_k = \chi \circ \exp(j\Psi_k)$ on $\Omega_3$. By the definition of $\Omega_3$, it suffices to show that

$$
|\chi(\xi^c) \exp(j\Psi_k(\xi^c)) - \chi(\xi^c)| \leq |\chi(\xi) \exp(j\Psi_k(\xi)) - \chi(\xi)|
$$

$$
\forall \xi \in I_k, \forall \xi^c \notin I_k. \quad (15)
$$

We first note that $\chi(\xi) = \chi(\xi^c)$ since $\chi$ is a uniform distribution. Condition (15) reduces to

$$
|\exp(j\Psi_k(\xi^c)) - 1| \leq |\exp(j\Psi_k(\xi)) - 1|
$$

$$
\forall \xi \in I_k, \forall \xi^c \notin I_k.
$$
By the definition of $I_k$, it holds that
\[
0 \leq |\Psi_k(\xi^c)| \leq |\Psi_k(\xi)| \leq \pi \quad \forall \xi \in I_k, \forall \xi^c \notin I_k.
\] (16)

Then we have for all $\xi \in I_k$ and $\xi^c \notin I_k$ that
\[
|\exp(j\Psi_k(\xi^c)) - 1| = \sqrt{(\cos \Psi_k(\xi^c) - 1)^2 + \sin^2 \Psi_k(\xi^c)} \\
= \sqrt{2 - 2 \cos |\Psi_k(\xi^c)|} \\
\leq \sqrt{2 - 2 \cos |\Psi_k(\xi)|} \\
= |\exp(j\Psi_k(\xi)) - 1|,
\] (17)
where the only inequality follows from (16). Hence we have proved that $x_{k+1} \in P_{\Omega_3}(z_k)$ and the proof is complete. □

Lemma 6.1 allows us to fully apply the convergence theory for the cyclic projections algorithm developed in [25, Section 3.1] to the SOPR algorithm. We next present an amount of mathematical material which is sufficient for formulating a local linear convergence criterion in a concise and memorable statement. Since phase retrieval is a typical nonconvex problem and our approach requires no convex relaxations, one should not expect global convergence for SOPR.

Recall that a set $\Omega$ is called prox-regular at a point $x^* \in \Omega$ if the projector $P_{\Omega}$ is single-valued around $x^*$ [30]. The analysis regarding prox-regularity in the context of the phase retrieval problem was first given in [23]. The following lemma provides the geometry of the auxiliary feasibility problem under consideration:

\[
\text{find } x \in \Omega_1 \cap \Omega_2 \cap \Omega_3.
\] (18)

**Lemma 6.2.** The following statements regarding the sets $\Omega_i$ $(i = 1, 2, 3)$ hold true.

(i) The set $\Omega_1$ is prox-regular at every point of it.

(ii) The set $\Omega_2$ is prox-regular at every point of it.

(iii) If $\hat{s}$ is the sparsity of the solutions with sparsest phase to problem (18), then the set $\Omega_3$ is prox-regular at every point $x^* \in \Omega_1 \cap \Omega_2 \cap \Omega_3$.

As a consequence, in the setting of item (iii), the sets $\Omega_i$ $(i = 1, 2, 3)$ are prox-regular at every solution to problem (18).

**Proof.** (i) The proof follows the idea of [23, Section 3.1]. We first note that the set $C$ defined at (10) is prox-regular at every point of it since it is the cartesian product of a number of circles which are typical examples of prox-regularity. It is clear from the definition of $\Omega_1$ that $\Omega_1 = D^{-1}(C)$ with $D$ given by (13). But $D^{-1}$ which is a unitary transform does not affect the geometry properties of $C$ and hence the statement is proved.

(ii) The first argument for item (i) also encompasses this statement.

(iii) We first note that $\Omega_3$ is the union of $\hat{s}$ copies of the circle $\{c \in \mathbb{C} \mid |c| = \chi\}$ and $n^2 - \hat{s}$ singletons $\{\}$.

Note that each of the components of $\Omega_3$ is a prox-regular set by the same argument as for Lemma 6.2 (i). Let $x^* \in \Omega_1 \cap \Omega_2 \cap \Omega_3$. By the definition of the sets $\Omega_i$ $(i = 1, 2, 3)$, we have $x^* = \chi \circ \exp(j \Phi^*)$ and $\|\Phi^*\|= \hat{s}$. Let us denote $I_{x^*}$ the set of indices corresponding to nonzero entries of $\Phi^*$. Since $\hat{s}$ is the sparsity of the solutions with sparsest phase to problem (18), there is the unique component of $\Omega_3$ which contains $x^*$, denoted $L_{x^*}$. Since $L_{x^*}$ is prox-regular as mentioned above, all we need is to show the existence of a neighborhood $U$ of $x^*$ such that
\[
\Omega_3 \cap U = L_{x^*} \cap U.
\] (19)

Let us define
\[
\theta := \min \{|\Phi^*(\xi)| : \xi \in I_{x^*}\} \in (0, \pi],
\] (20)
\[
\delta := \chi \sqrt{2 - 2 \cos \theta} > 0.
\] (21)
We will show that the neighborhood
\[ U := \{ x \in \mathbb{C}^{n \times n} \mid \| x - x^* \|_1 < \delta \} \]  
(22)
of \( x^* \) satisfies the desired property (19). Since the inclusion \((\Omega_3 \cap U) \supset (L_{x^*} \cap U)\) is trivial, we now prove only the inverse one, \((\Omega_3 \cap U) \subset (L_{x^*} \cap U)\). Indeed, take any \( x \in \Omega_3 \cap U \). Note that \( |x| = \chi \) as \( x \in \Omega_3 \). Let us write \( x = \chi \odot \exp (j\Phi) \). Then for every index \( \xi \in \mathcal{I}_{x^*} \), we have by the triangle inequality, (17), (22), (20) and (21) successively that
\[ |x(\xi) - \chi| \geq |x^*(\xi) - \chi| - |x^*(\xi) - x(\xi)| \]
\[ = \chi \sqrt{2 - 2 \cos \Phi^*(\xi)} - |x^*(\xi) - x(\xi)| \]
\[ > \chi \sqrt{2 - 2 \cos \Phi^*(\xi)} - \delta \]
\[ \geq \chi \sqrt{2 - 2 \cos \theta} - \delta \]
\[ = \delta - \delta = 0. \]
This in particular implies that \( \Phi(\xi) \neq 0 \) for all indices \( \xi \in \mathcal{I}_{x^*} \). \( \| \Phi \|_0 = \hat{s} \). That is \( x \in L_{x^*} \) and hence the proof is complete.

It is worth mentioning that the condition on the uniform distribution of the amplitude \( \chi \) is essential to both Lemmas 6.1 and 6.2.

We are now ready to state and prove a local linear convergence criterion for SOPR, which has been known from Lemma 6.1 to be the same as the cyclic projections algorithm \( T_{\text{CP}} := P_{\Omega_1}P_{\Omega_2}P_{\Omega_3}. \)

**Theorem 6.3 (R-linear convergence of SOPR).** Let \( \Phi^* \) be a sparsest phase solution to SPR problem (2) with \( \| \Phi^* \|_0 = \hat{s} \). Suppose that the set-valued mapping \( F := \text{Id} - T_{\text{CP}} \) is metrically subregular at \( x^* := \chi \odot \exp (j\Phi^*) \) for 0, i.e., there is a constant \( \kappa > 0 \) such that the inequality
\[ \kappa \text{dist}(x, F^{-1}(0)) \leq \text{dist}(0, F(x)) \]
holds for all \( x \) in a neighborhood of \( x^* \). Then every iterative sequence generated by SOPR converges R-linearly to a solution to SPR provided that the initial point is sufficiently close to \( x^* \).

**Proof.** Since \( \hat{s} \) is the sparsity of the sparsest phase solutions to the SPR problem (2), \( \Phi^* \) is a sparsest phase solution to (2) if and only if \( x^* = \chi \odot \exp (j\Phi^*) \) is a solution to (18). By Lemma 6.1 the two algorithms SOPR and \( T_{\text{CP}} \) coincide. Therefore, it suffices to show the conclusion of Theorem 6.3 for sequences generated by \( T_{\text{CP}} \). By Lemma 6.2 the sets \( \Omega_i \) \((i = 1, 2, 3)\) are prox-regular at \( x^* \). By [14, Theorem 2.14 (ii)], the projections \( P_{\Omega_i} \) \((i = 1, 2, 3)\) are almost firmly nonexpansive with violation that can be made arbitrarily small by shrinking the effective neighborhood of \( x^* \). Then by [25, Proposition 2.4 (iii)], there is a neighborhood \( U \) of \( x^* \) such that the operator \( T_{\text{CP}} \) is almost averaged on \( U \) with averaging constant 3/4 and violation \( \varepsilon \) that can be made arbitrarily small. Hence, taking also the metric subregularity of \( \text{Id} - T_{\text{CP}} \) at \( x^* \) for 0 into account, we can directly apply [25, Corollary 2.3] to obtain local linear convergence of \( T_{\text{CP}} \). The proof is complete.

The notion of metric subregularity appearing in Theorem 6.3 is amongst cornerstones of variational analysis and optimization theory with many important applications [9,16]. In the context of feasibility problem, the metric subregularity is closely related to the mutual arrangement of the sets at the reference point [14,17-21,25]. A deeper look at this type of regularity for the collection of sets \( \{\Omega_1, \Omega_2, \Omega_3\} \) at \( x^* \) involved in Theorem 6.3 is obviously of importance but beyond the scope of this paper.

### 7 Numerical experiments

Several numerical aspects of SOPR including solvability, sensitivity to sparsity parameter, sensitivity to erroneous input and robustness against noise are examined on relevant experiments. Since algorithms for retrieving sparse signal such as those suggested in [28,31] are not applicable to SPR as discussed in Section 3, we cannot make a comparison of SOPR to them. Indeed, we have not found in the literature any numerical experiments on (locally) solving sparse phase retrieval given only one PSF intensity image. For comparison, we primarily compare SOPR to the well known Gerchberg-Saxton (GS) algorithm, which
mathematically corresponds to SOPR with \( \hat{s} = n^2 \) (i.e., steps (6) and (7) of SOPR are dispensed). The ultimate conclusion to be made is that SOPR, which can be thought as a \( \| \cdot \|_q \)-regularization variant of GS works significantly better than that algorithm for SPR. This advantageous feature of SOPR is also demonstrated for the easier problem, SPR with multiple images, in Section 7.5.

During the discussion regarding the sparsity level \( s \) of the signal and the sparsity parameter \( \hat{s} \), we also compare SOPR to the sparse Fienup (SF) algorithm [26] applied to retrieve the sparse signal \( x - \chi \) from the data \( |F(x - \chi)| \) which is approximately calculated from \( |F(x)| \) via Lemma 4.1. For convenience, we also specify SF for our SPR problem in Algorithm 7.1. It is clear that the inexactness of the calculation of \( b' \) from \( b \) and \( \chi \) described in Algorithm 7.1 is an obvious factor leading to the ineffectiveness of this algorithm for solving SPR as will be observed in this section. It is worth mentioning that the outperform of SOPR over SF is not the main goal of these experiments but rather the performance of SOPR itself.

**Algorithm 7.1. [SF for SPR]**

**Input:**
- \( b \in \mathbb{R}_{+}^{n \times n} \) — intensity PSF image
- \( \hat{s} \) — sparsity parameter
- \( \Phi_0 \) — initial guess for \( \Phi \)
- \( \tau \) — tolerance threshold.

**Data processing:**
- calculate \( b' \) from \( b \) and \( \chi \) using Lemma 4.1

\[ \chi = \chi \odot \exp(j \Phi_0) - 1 \]  — initial guess of the sparse signal \( x' = x - \chi \).

**Iteration process:** given \( u_k \)

1. \( U_k = F(u_k) \) — Fourier transform
2. \( V_k = \sqrt{b'} \odot \frac{U_k}{|U_k|} \) — amplitude constraint
3. \( v_k = F^{-1}(V_k) \) — inverse Fourier transform
4. find \( I_k \), a set of indices corresponding to \( n^2 - \hat{s} \) smallest entries of \( v_k \)
5. \( u_{k+1} = v_k \) on \( I_k \) and zero elsewhere.

**Stopping criterion:** \( \|\Phi_k - \Phi_{k+1}\| < \tau \) or \( k \) exceeds a number of iterates, where \( \Phi_k := \arg(u_k + \chi) \).

**Output:** \( \hat{\Phi} = \arg(u_{\text{end}} + \chi) \) — the estimate.

### 7.1 Solvability Versus Sparsity

In this section we examine the recovery probability of SOPR relative to the sparsity of the phase. A comparison to SF is also presented. For each sparsity level \( s \), an \( s \)-sparse phase \( \Phi \) of size \( 32 \times 32 \) pixels was randomly generated from the uniform distribution on \( [-\pi, \pi] \). For this experiment very little white Gaussian noise, generated by using the Matlab function awgn(\( ., 40dB; \)‘measured’), was added to the squared amplitude of \( F(\chi \odot \exp(j \Phi)) \), the negative entries of the obtained measurement were then reset to zeros. In the remainder of this paper, the amplitude \( \chi \) is uniform with value 1. One hundred initial points were generated according to the formula \( x_0 = \chi \odot \exp(j (\Phi + \phi)) \), where the entries of \( \phi \) were randomly generated from the uniform distribution on \( [-2\pi/5, 2\pi/5] \). SOPR and SF were called 100 times for each sparsity level \( s = 4 + 10k \) with \( k = 1, 2, \ldots, 58 \), respectively. The sparsity parameter \( \hat{s} \) was taken around 105% of \( s \). In this experiment and in the sequel, a run is counted as successful if the criterion

\[
\|\hat{\Phi} - \Phi\|_1 \leq \tau := s 10^{-2}
\]

is reached in less than 200 iterates, where \( \hat{\Phi} \) is the truncated at \( s \) largest entries of the current estimated phase. Fig. 1 shows that the recovery probability of SOPR is almost surely for the sparsity between 6% and 25%. The algorithm fails to retrieve the phase in both extreme cases of sparsity: lower 4% or higher 40%. It is also clear that SF is far outperformed by SOPR.
7.2 Sensitivity to Sparsity Parameter

Likewise the stability of projection methods for solving *sparse affine feasibility* with respect to the sparsity parameter [15], in this section we demonstrate that SOPR for locally solving sparse phase retrieval problems also enjoys that kind of stability. A comparison to SF is presented. We consider a simulation sparse phase $\Phi$ of size $32 \times 32$, intensity image corrupted by white Gaussian noise at $40dB$, 100 initial points and the criterion of success run (23) as in Section 7.1. In this experiment we know the sparsity of $\Phi$ being $s = 144$ and examine the recovery probability of SOPR for a wide range of sparsity parameter $\hat{s}$, both lower and higher than $s$. SOPR and SF were called 100 times for each sparsity parameter $\hat{s} = 44 + 10k$ with $k = 1, 2, \ldots, 50$, respectively. It has been known from Fig. 1 that the recovery probability of SOPR with $\hat{s}$ slightly above $s$ is almost surely. Fig. 2 shows that such a probability indeed remains stable for a wide range of $\hat{s}$ - between 80% and 180% of $s$. It is also clear from Fig. 2 that SF is far outperformed by SOPR.

Figure 1: Recovery probability versus sparsity of SOPR and SF.

Figure 2: Sensitivity to sparsity parameter of SOPR and SF.
7.3 Sensitivity to Erroneous Input

In this section we examine the sensitivity of SOPR with respect to the erroneous input, the distance from the initial point to the solution. A comparison to SF and the Gerchberg-Saxton (GS) methods is presented. We consider a simulation phase $\Phi$ of size $32 \times 32$ with sparsity $s = 144$, a white Gaussian noise and the success criterion (23) as in Section 7.2. The sparsity parameter was taken around 105% of $s$. Note that for an initial phase guess $\Phi_0 = \Phi + \phi$, due to the sparseness of $\Phi$ a norm of $\phi$ might be not very informative to the intrinsic difference between $\Phi$ and $\Phi_0$. For example, any $\phi$ that if was added to $\Phi$ would have no affects on the set of indices of $s$ largest entries of $\Phi$ could be considered as zero in this experiment. Here, the erroneous input is measured as the difference (in pixels) between the supports of $\Phi$ and $\Phi_0$ the truncation of $\Phi_0$ at $s$-largest entries. The minimum of such a support difference is zero pixel corresponding to the ignorable case of $\phi$ described above, and its maximum value is $2s$ corresponding to the opposite scenario. In this experiment, 100 initial points were generated for each support difference value $d = 5k$ with $k = 0, 1, \ldots, 40$, respectively. The three algorithms SOPR, SF and GS were called 100 times for those initial points. Fig. 3 shows that SOPR performs consistently for $d$ up to 85% of $s$. SF performs well for $d$ up to 50% of $s$ while GS can not retrieve the phase for $d$ above 15% of $s$. It is clear that SOPR outperforms SF and GS in this experiment. The outperform of SOPR over GS in particular demonstrates the effectiveness of our regularization scheme newly proposed in this paper.

7.4 Robustness Against Noise

In this section we examine the robustness of SOPR against noise. We consider a simulation phase $\Phi$ of size $32 \times 32$ with sparsity $s = 144$ and the success criterion (23) as in Section 7.2. The sparsity parameter was taken around 105% of $s$. One hundred initial points were generated according to the formula $x_0 = \chi \odot \exp(j(\Phi + \phi))$, where $\phi$ was randomly generated from the uniform distribution on $[-2\pi/5, 2\pi/5]$ as in Section 7.1. The support difference is on average about 117 pixels (see Section 7.3). Since the Gerchberg-Saxton method does not work in this experiment as can be observed from Fig. 3, we report the performance of SOPR and SF. For each level of white Gaussian noise ranging from $20dB$ to $50dB$, SOPR and SF were called 100 times with the initial points described above. Fig. 4 shows that for signal-to-noise rate below 26$dB$ they can not retrieve the phase and that for signal-to-noise rate above 30$dB$ SOPR again far outperforms SF.
In this section we briefly compare the cyclic projections (CP) algorithm with the corresponding cyclic version of SOPR for SPR given multiple images. We consider a simulation phase $\Phi$ of size $128 \times 128$ with sparsity level 25%, which was randomly generated from the uniform distribution on $[-\pi, \pi]$. A circular aperture with radius 64 pixels was used in this experiment. Three intensity images corresponding to the three phase diversity patterns $\exp\left(j\pi d Z_0^2(\rho)\right)$ with $d = -2, 0, 3$ respectively were generated, where $Z_0^2(\rho)$ is the radial polynomial in the polar coordinates $(\rho, \varphi)$. For the experiment with noise, a white Gaussian noise at 30$\text{dB}$ was added to these images and the negative entries of the obtained images were then reset to zeros. The sparsity parameter for (the cyclic version of) SOPR was taken about 105% of the true sparsity. The amplitude distribution (zero phase everywhere) was taken as the starting point for both algorithms. Note that a good initial point is not technically required for phase retrieval given multiple images. For this simulation we can actually monitor the residual between the estimated phase $\hat{\Phi}$ and the exact phase $\Phi$, and hence the residual quantity $\|\hat{\Phi} - \Phi\|_F$ will be used to evaluate the quality of retrieval in this section. In the noise-free case shown by Fig. 5 (left), SOPR appears to recover the exact phase while the cyclic projections algorithm seemingly fails to exactly retrieve the zero entries of the true solution. In the experiment with noise shown by Fig. 5 (right), SOPR also appears to yield the estimated phase with smaller residual than that of CP.

8 Conclusion

We have developed a new $\| \cdot \|_0$-regularization scheme which is implicitly incorporated in the SOPR algorithm for solving the sparse phase retrieval (SPR) problem. For each iterate of SOPR, apart from the well known steps for Fourier transform-based phase retrieval algorithms, an additional sparsity constraint is applied to the phase to capture its sparsity property. This additional step is mathematically a rotation
on the underlying space $\mathbb{C}^{n \times n}$ but with direction varying in iterates. In the setting of SPR with a uniform amplitude distribution, we have proved that this step indeed corresponds to a projection on an auxiliary set independent of the iterates. This result is rather unexpected since the $\| \cdot \|_0$-regularization applied to the phase on $(-\pi, \pi)^{n \times n}$ turns out to be equivalent to a metric projection on the complex Hilbert space $\mathbb{C}^{n \times n}$. As a result, SOPR algorithm is shown to be equivalent to the cyclic projections algorithm for solving a feasibility problem involving several auxiliary locally prox-regular sets. The convergence analysis of SOPR is rigorously obtained based on recent results for the cyclic projections algorithm for feasibility. Our numerical results have demonstrated clear empirical advantages of our $\| \cdot \|_0$-regularization scheme. The overall message to deliver is that if the phase is known a priori to be sparse, then a sparsity constraint appropriately applied to the phase will improve the recovery quality. For a concrete example, we have illustrated those advantages on the SPR problem given only one PSF intensity image. To this end the primary research goal of this paper has been achieved.

To additionally demonstrate the potential of our approach for more application settings, we have also preliminarily shown the seemingly advantage of our strategy for the sparse phase retrieval problem with multiple intensity images. Extensive numerical comparison to existing algorithms in this setting is vast and will be postponed for a future research.

Acknowledgment

The research leading to these results has received funding from the European Research Council under the European Union’s Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement No. 339681. DRL was supported in part by German Israeli Foundation Grant G-1253-304.6 and Deutsche Forschungsgemeinschaft Collaborative Research Center SFB755.

References

[1] J. Antonello and M. Verhaegen, “Modal-based phase retrieval for adaptive optics,” J. Opt. Soc. Am. A, vol. 32, no. 6, pp. 1160–1170, 2015.

[2] 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. Amer. A, vol. 19, no. 7, pp. 1334–1345, 2002.

[3] H. H. Bauschke, D. R. Luke, M. Hung Phan, and X. Wang, “Restricted normal cones and sparsity optimization with affine constraints,” Found. Comput. Math., vol. 14, no. 1, pp. 63–83, 2014.

[4] M. R. Bolcar, “Phase Diversity for Segmented and Multi-aperture Systems,” University of Rochester, New York, 2008.

[5] J. M. Borwein and M. K. Tam, “A cyclic Douglas-Rachford iteration scheme,” J. Optim. Theory Appl., vol. 160, no. 1, pp. 1–29, 2014.

[6] J. J. M. Braat, P. Dirksen, A. J. E. M. Janssen, S. van Haver, and A. S. van de Nes, “Extended Nijboer-Zernike approach to aberration and birefringence retrieval in a high-numerical-aperture optical system,” J. Opt. Soc. Am. A., vol. 22, no. 12, pp. 2635–2650, 2005.

[7] E. Brunner, C. C. de Visser, and M. Verhaegen, “Nonlinear spline wavefront reconstruction from Shack-Hartmann intensity measurements through small aberration approximations,” J. Opt. Soc. Am. A., vol. 34, no. 9, pp. 1535–1549, 2017.

[8] R. Doelman, H. Thao Nguyen, and M. Verhaegen, “Solving large-scale general phase retrieval problems via a sequence of convex relaxations,” arxiv.org/abs/1803.02652, 2018.

[9] A. L. Dontchev and R. T. Rockafellar, “Implicit Functions and Solution Mappings,” Springer-Verlag, New York, second edition, 2014.

[10] J. R. Fienup, “Phase retrieval algorithms: a comparison,” Appl. Opt., vol. 21, pp. 2758–2769, 1982.
[11] J. R. Fienup, “Phase retrieval algorithms: a personal tour,” Appl. Opt., vol. 52, no. 1, pp. 45–56, 2013.
[12] R. W. Gerchberg and W. O. Saxton, Optik (Stuttgart), 35–237, 1972.
[13] J. W. Goodman, “Introduction to Fourier Optics,” Roberts & Company Publishers, 2005.
[14] R. Hesse and D. R. Luke, “Nonconvex notions of regularity and convergence of fundamental algorithms for feasibility problems,” SIAM J. Optim., vol. 23, no. 4, pp. 2397–2419, 2013.
[15] R. Hesse, D. R. Luke, and P. Neumann, “Alternating projections and Douglas-Rachford for sparse affine feasibility,” IEEE Trans. Signal Process., vol. 62, no. 18, pp. 4868–4881, 2014.
[16] D. Klatte and B. Kummer, “Nonsmooth Equations in Optimization,” Kluwer Academic Publishers, Dordrecht, 2002.
[17] A. Y. Kruger, “About regularity of collections of sets,” Set-Valued Anal., vol. 14, no. 2, pp. 187–206, 2006.
[18] A. Y. Kruger, “About stationarity and regularity in variational analysis,” Taiwanese J. Math., vol. 13, no. 6A, pp. 1737–1785, 2009.
[19] A. Y. Kruger, D. R. Luke, and H. Thao Nguyen, “Set regularities and feasibility problems,” Math. Program., vol. 168, no. 1-2, Ser. B, pp. 279–311, 2018.
[20] A. Y. Kruger, D. R. Luke, and H. Thao Nguyen, “About subtransversality of collections of sets,” Set-Valued Var. Anal., vol. 25, no. 4, pp. 701–729, 2017.
[21] A. Y. Kruger and H. Thao Nguyen, “Quantitative characterizations of regularity properties of collections of sets,” J. Optim. Theory and Appl., vol. 164, no. 1, pp. 41–67, 2015.
[22] D. R. Luke, “Relaxed averaged alternating reflections for diffraction imaging,” Inverse Problems, vol. 15, no. 21, pp. 37–50, 2005.
[23] D. R. Luke, “Finding best approximation pairs relative to a convex and prox-regular set in a Hilbert space,” SIAM J. Optim., vol. 19, no. 2, pp. 714–739, 2008.
[24] D. R. Luke, J. V. Burke, and R. G. Lyon, “Optical wavefront reconstruction: theory and numerical methods,” SIAM Rev., vol. 44, no. 2, pp. 169–224, 2002.
[25] D. R. Luke, H. Thao Nguyen, and M. K. Tam, “Quantitative convergence analysis of iterated expansive, set-valued mappings,” Math. Oper. Res., accepted.
[26] S. Mukherjee and C. Seelamantula, “An iterative algorithm for phase retrieval with sparsity constraints: Application to frequency domain optical coherence tomography,” Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., pp. 553–556, 2012.
[27] H. Thao Nguyen, “A convergent relaxation of the Douglas-Rachford algorithm,” Comput. Optim. Appl., doi.org/10.1007/s10589-018-9989-y.
[28] E. J. R. Pauwels, A. Beck, Y. C. Eldar, and S. Sabach, “On Fienup methods for sparse phase retrieval,” IEEE Trans. Signal Process., vol. 66, no. 4, 2018.
[29] P. J. Piscaer, A. Gupta, O. Soloviev, and M. Verhaegen, “Modal-based phase retrieval using Gaussian radial basis functions,” submitted.
[30] R. A. Poliquin, R. T. Rockafellar, and L. Thibault, “Local differentiability of distance functions,” Trans. Amer. Math. Soc., vol. 352, no. 11, pp. 5231–5249, 2000.
[31] Y. Shechtman, A. Beck, and Y. C. Eldar, “GESPAR: efficient phase retrieval of sparse signals,” IEEE Trans. Signal Process., vol. 62, no. 4, pp. 928–938, 2014.
[32] 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*, vol. 32, no. 3, pp. 87–109, 2015.

[33] W. H. Southwell, “Wave-front analyzer using a maximum likelihood algorithm,” *J. Opt. Soc. Am.*, vol. 67, no. 3, pp. 396–399, 1977.