Proximity Operators for Phase Retrieval

Ferréol Soulez, Éric Thiébaut, Antony Schutz, André Ferrari, Frédéric Courbin, Michael Unser

Proximity Operators for Phase Retrieval

FERREOL SOULEZ[1,*], ÉRIC THIEBAUT[2], ANTONY SCHUTZ[3], ANDRE FERRARI[3], FREDEIC COURBIN[4], AND MICHAEL UNSER[1]

[1] Biomedical Imaging Group, École polytechnique federale de Lausanne (EPFL), Lausanne CH-1015, Switzerland.
[2] Univ Lyon, Univ Lyon1, Ens de Lyon, CNRS, Centre de Recherche Astrophysique de Lyon UMR5574, F-69230, Saint-Genis-Laval, France
[3] Lab. J.-L. Lagrange, Universite de Nice Sophia Antipolis, CNRS, Observatoire de la Côte d'Azur, Parc Valrose, F-06108 Nice cedex 02, France
[4] Laboratoire d'astrophysique, École Polytechnique Federale de Lausanne (EPFL), Observatoire de Sauverny CH-1290 Versoix, Switzerland.
[*] Corresponding author: ferreol.soulez@epfl.ch

Compiled November 13, 2017

We present a new formulation of a family of proximity operators that generalize the projector step for phase retrieval. These proximity operators for noisy intensity measurements can replace the classical "noise free" projection in any projection-based algorithm. They are derived from a *maximum likelihood* formulation and admit closed form solutions for both the Gaussian and the Poisson cases. In addition, we extend these proximity operators to undersampled intensity measurements. To assess their performance, these operators are exploited in a classical Gerchberg Saxton algorithm. We present numerical experiments showing that the reconstructed complex amplitudes with these proximity operators perform always better than using the classical intensity projector while their computational overhead is moderate.

OCIS codes: 100.5070, 100.3020, 100.3190,100.6640

http://dx.doi.org/10.1364/ao.XX.XXXXXX

1. INTRODUCTION

The classical phase-retrieval problem is to reconstruct a complex-valued signal $\boldsymbol{x}$ from measurements of its squared modulus [1]. This problem arises in many applications (*e.g.*, cristallography [2], microscopy [3], astronomy [4]). Since the seminal paper of Gerchberg and Saxton [5], an abundant literature has been devoted to it (see [6] for a review). A large part of the proposed algorithms relies on successive projections [7–11]. For the last few years, there is a renewed interest for phase retrieval and several new alternatives to successive projections methods have been proposed: the semi-definite-programming based formulations [12, 13], the algorithms for phase retrieval of sparse signal [14–16] the gradient based methods using Wirtinger derivatives [17, 18] and a variational Bayesian framework [19].

Here, we adopt a vector representation of the complex image $\boldsymbol{x} = (x_1, \ldots, x_K)$, where K is the number of pixels. In the phase retrieval problem, the forward model that links the complex amplitude $\boldsymbol{x} \in \mathbb{C}^K$ to the measured image intensities $\boldsymbol{d} \in \mathbb{R}_+^K$ is

$$d_k = |x_k|^2 + n_k\,, \tag{1}$$

where $\boldsymbol{n}$ is some measurement noise and $|x_k|^2$ denotes the squared modulus of x_k.

Such an inverse problem is classically solved in a variational framework by estimating $\boldsymbol{x}$ that minimizes a cost function

$$\mathcal{C}(\boldsymbol{x}) = \mathcal{L}(\boldsymbol{x}) + \mathcal{R}(\boldsymbol{x})\,, \tag{2}$$

which is a sum of the data term $\mathcal{L}$ and a regularization functional $\mathcal{R}$. In this approach known as penalized *maximum likelihood*, the data term is defined according to the forward model and the statistics of the noise, whereas the regularization function is designed to enforce some prior knowledge about $\boldsymbol{x}$ (such as support, non-negativity, smoothness,...). As $\mathcal{L}$ and $\mathcal{R}$ are defined independently, any improvement on one of these functions implies a better estimate of the solution of the inverse problem.

Most projection-based algorithms [5, 7–11] use constraints that assume noise-free measurements. Some authors have studied the behavior of these methods in noisy environment [20] while others have proposed empirical modifications to mitigate the effect of the noise [21–23]. In this paper, we derive a likelihood function adapted to the statistics of the noise via a simple modification of the intensity-projection operator. We had previously established the formulation of this proximity operator in the Gaussian case with a specific ADMM algorithm for image reconstruction in optical long-baseline interferometry [24] ; a similar result was also published recently [25]. But neither further characterization nor comparison with standard projection methods were done.

Rather than a full-fledged phase-retrieval algorithm, the scope of this paper is a novel formulation of a noise-adapted projection step that can be used in any other projection-based algorithm [8]. Therefore, we focus on the likelihood function. To demonstrate its effect, we apply the proposed projectors in the standard Gerchberg Saxton algorithm (GS). We have chosen this simple phase-retrieval algorithm as a baseline because it does not use any priors. Therefore, the quality of its results depend only on the projection used. Whereas GS is hardly state-of-the-art anymore, the reader must keep in mind that our proposed proximity operators can be plugged into many optimization schemes (see [10, 26]) that rely on proximity operators to minimize a regularized cost function.

2. GERCHBERG-SAXTON ALGORITHM

The error-reduction method (GS), described in Algorithm (1), estimates the complex amplitude (the wavefront) of a light wave in the plane z_A from the intensity profiles $\boldsymbol{d}_A$ and $\boldsymbol{d}_B$ measured at depth z_A and z_B, respectively. It solves the feasibility problem

$$\text{find } \boldsymbol{x} \in C_A \cap \{\boldsymbol{x} : \mathbf{H}\,\boldsymbol{x} \in C_B\}\ , \tag{3}$$

where $\mathbf{H}$ is the propagation operator from plane z_A to plane z_B and C_i with $i = A, B$ is the set of complex-valued signals of squared modulus $\boldsymbol{d}_i$, *i.e.* $C_i = \{\boldsymbol{x} \in \mathbb{C}^K, |\boldsymbol{x}|^2 = \boldsymbol{d}_i\}$. The propagation operator is classically either the Fresnel operator (under a Fresnel approximation) or the Fourier operator (under a Fraunhoffer approximation). This can be reformulated as the minimization problem

$$\boldsymbol{x}^+ \in \underset{\boldsymbol{x}\in\mathbb{C}^K}{\arg\min}\left(\imath_{C_A}(\boldsymbol{x}) + \imath_{C_B}(\mathbf{H}\,\boldsymbol{x})\right)\ , \tag{4}$$

where $\imath_C$ is the indicator function of the set C defined as

$$\imath_C(\boldsymbol{x}) = \begin{cases} 0\,, & \text{if } \boldsymbol{x} \in C \\ +\infty\,, & \text{otherwise.} \end{cases} \tag{5}$$

Observe that, when $\boldsymbol{d}_i \neq \mathbf{0}$, C_i is generally not a convex set. Therefore, only local convergence can be established [27].

Algorithm 1. Gerchberg-Saxton algorithm

1: **procedure** GS($\boldsymbol{d}_A, \boldsymbol{d}_B$)
2: $\quad \boldsymbol{x}^{(0)} = \sqrt{\boldsymbol{d}_A}$ ▷ Initialization
3: $\quad$ **for** $n = 1, 2, \ldots,$ maxiter **do**
4: $\quad\quad \boldsymbol{y}^{(n-1/2)} = \mathbf{H}\cdot\boldsymbol{x}^{(n-1)}$ ▷ Propagation to the z_B plane
5: $\quad\quad \boldsymbol{y}^{(n)} = P_B(\boldsymbol{y}^{(n-1/2)})$ ▷ Projection
6: $\quad\quad \boldsymbol{x}^{(n-1/2)} = \mathbf{H}^{-1}\cdot\boldsymbol{y}^{(n)}$ ▷ Back propagation to the z_A plane
7: $\quad\quad \boldsymbol{x}^{(n)} = P_A(\boldsymbol{x}^{(n-1/2)})$ ▷ Projection
8: $\quad$ **return** $\boldsymbol{x}^{(\text{maxiter})}$ ▷ The complex amplitude in the z_A plane

The GS algorithm and its successors [8, 10, 11] involve an element-wise projection operator $P(\boldsymbol{x}\,|\,\boldsymbol{d}) = (P(x_1\,|d_1), \ldots, P(x_K\,|d_K))$ that constrains the modulus of the current iterate $\boldsymbol{x}$ to be equal to the square root of its measurement $\sqrt{\boldsymbol{d}}$ while keeping its phase untouched, as in

$$P(x_k\,|d_k) = \begin{cases} \frac{x_k}{|x_k|}\sqrt{d_k}\,, & \text{if } |x_k| > 0 \\ \sqrt{d_k}\,, & \text{otherwise.} \end{cases} \tag{6}$$

The projection $P(\boldsymbol{x}\,|\,\boldsymbol{d})$ of $\boldsymbol{x}$ onto the set C of all signals of intensity (or squared modulus) $\boldsymbol{d}$ will be called "classical projection" throughout this paper. It is a solution of

$$\underset{\boldsymbol{y}\in\mathbb{C}^N}{\text{minimize}}\left(\imath_C(\boldsymbol{y}) + \frac{1}{2}\,\|\boldsymbol{x}-\boldsymbol{y}\|^2\right)\ . \tag{7}$$

To prevent stagnation of the GS algorithm, a relaxed projection step P' was proposed [28, 29]:

$$P'(x_k\,|d_k) = (1-\beta)\,x_k + \beta\,P(x_k\,|d_k)\,, \tag{8}$$

where $0 \leq \beta \leq 1$ is a relaxation parameter empirically set close to 0 for regions where the noise dominates.

As observed by Levi and Stark [30, 31], the GS algorithm is a non-convex instance of the projection-onto-convex-set (POCS) algorithm. POCS is widely employed in signal processing to solve feasibility problems. However, as soon as noisy intensities are considered, equation (6) does not anymore give the solution that is optimal in the maximum-likelihood sense. Therefore, GS leads to errors in the reconstructed wavefront in the presence of noisy measurements.

We assume that the measurement noise $n_k = d_k - |x_k|^2$ at pixel k is independent and centered with a probability density $\Pr(n_k\,|\,x_k)$. For a given intensity measurement d_k, the co-log-likelihood of the noise distribution at pixel k (up to the constant cst) is:

$$\ell_k(n_k) = -\ln\Pr(n_k\,|\,x_k) + \text{cst}\,. \tag{9}$$

The problem addressed by GS has a maximum-likelihood formulation expressed by

$$\boldsymbol{x}^+ \in \underset{\boldsymbol{x}\in\mathbb{C}^K}{\arg\min}\left(\sum_{k=1}^{K}\ell_k\left(|x_k|^2 - d_k\right) + \sum_{k'=1}^{K}\ell_{k'}\left(|[\mathbf{H}\,\boldsymbol{x}]_{k'}|^2 - d_{k'}\right)\right)\ . \tag{10}$$

This is not a feasibility problem anymore. However, it is still closely related to the GS formulation described by Equation (3). We argue that, with the help of proximal operators, both problems can be solved using identical convex-optimization techniques (*e.g.*, Douglas-Rachford) without relying on smooth approximations of ℓ [32].

3. PROXIMITY OPERATOR FOR INTENSITY

A. Non-Convex Proximity Operators

It is possible to tackle a class of problems broader than feasibility problems by introducing proximity operators [26]. A proximity operator (or Moreau proximal mapping [33]) is a generalization of the classical projection on a set where the indicator function $\imath_C$ in (7) is replaced by an arbitrary lower semi-continuous convex function $g : \mathbb{C}^K \to \mathbb{R}$ so that

$$\text{prox}_g(\widetilde{\boldsymbol{x}}) \stackrel{\text{def}}{=} \underset{\boldsymbol{x}\in\mathbb{C}^K}{\arg\min}\left(g(\boldsymbol{x}) + \frac{1}{2}\,\|\boldsymbol{x}-\widetilde{\boldsymbol{x}}\|_2^2\right)\ . \tag{11}$$

The concept of proximal mapping has also been extended to non-convex functions that fulfill three conditions: (i) lower semi-continuity; (ii) *prox-boundedness*; and (iii) *prox-regularity* (see Theorem 4 of [34]).

B. Proximal Operator for Maximum-Likelihood

As long as the measurement noise is uncorrelated, the likelihood function defined in (9) is separable along pixels. In this element-wise operation, we shall drop the subscript k to simplify the notations and state $\alpha\,f(x) = \ell_k(d_k - |x_k|^2)$ with $\alpha > 0$ a tuning factor. The function f has the following properties: (i) *continuity*, provided that ℓ_k is also continuous (that is true for most noise statistics used in practice) ; (ii) *non convexity* (*e.g.*, if $x_1 = \sqrt{d}$ is a minimum of f, then $x_2 = -\sqrt{d}$ is, but not necessarily $(x_1 + x_2)/2$) ; (iii) *prox-boundedness* as f is positive (and

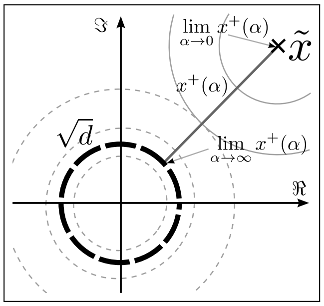

Fig. 1. Trajectory of $x^+(\alpha) = \mathrm{prox}_{\alpha f}(\tilde{x})$ as a function of α. $x^+(\alpha)$ follows the line where the level set of f (thin dashed circles) and $|x - \tilde{x}|^2$ (thin circles) are tangent.

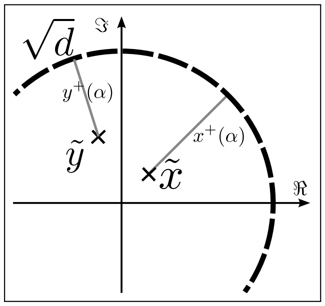

Fig. 2. Illustration of the expansiveness of $\mathrm{prox}_{\alpha f}(x)$ when $|x|^2 < d$
.

proper). However, as described further, f is not *prox-regular* in $x = 0$.

The proximity operator of $\alpha\, f$ is given by

$$\mathrm{prox}_{\alpha f}(\tilde{x}) = \underset{x\in\mathbb{C}}{\arg\min} \left\{ \alpha\, f(x) + \frac{1}{2}\, |x - \tilde{x}|^2 \right\}. \tag{12}$$

As f is a function that depends only on the squared modulus of x, the solution necessarily lies on the line passing through $\tilde{x}$ and 0 where the gradients of both parts of (12) have opposite directions. The phase of the solution is therefore the phase of $\tilde{x}$. The solution $x^+ = \rho^+ \exp(\jmath\, \phi^+)$ of (12) is given by

$$\begin{aligned} \rho^+ &= \underset{\rho\geq 0}{\arg\min} \left\{ \alpha\, f(\rho) + \frac{1}{2}\, (\rho - \tilde{\rho})^2 \right\}, & (13)\\ \phi^+ &= \tilde{\phi}, & (14) \end{aligned}$$

where $\tilde{x} = \tilde{\rho} \exp(\jmath\, \tilde{\phi})$.

Furthermore, if $f(x)$ has its minimum in $|x|^2 = d$ and $f(|x|)$ increases monotonically for $|x| > \sqrt{d}$, then there is a solution that lies on the line between $\tilde{x}$ and its projection on the circle $|x|^2 = d$, as illustrated in Figure 1. The position on this line varies monotonically with α, so that $\mathrm{prox}_{\alpha f}(\tilde{x})$ is $\tilde{x}$ for $\alpha = 0$ and gets closer to $\frac{\tilde{x}}{|\tilde{x}|} \sqrt{d_k}$ as α increases. The classical operator defined in (6) can thus be seen as $\lim_{\alpha\to\infty} \mathrm{prox}_{\alpha f}(\tilde{x}) = p(x_k\, |d_k)$. From this solution, we can identify three subdomains where $\mathrm{prox}_{\alpha f}(\tilde{x})$ has different properties.

- When $\tilde{x} \in \{x \in \mathbb{C}, |x|^2 \geq d\}$, $\mathrm{prox}_{\alpha f}(\tilde{x})$ is single valued and thus f is *prox-regular*. Furthermore, the proximity operator of f is non-expansive on this sub-domain.
- When $\tilde{x} \in \{x \in \mathbb{C}, 0 < |x|^2 < d\}$, f is still *prox-regular* but $\mathrm{prox}_{\alpha f}$ is no longer non-expansive. Indeed, $\|\, \mathrm{prox}_{\alpha f}(\tilde{x}) - \mathrm{prox}_{\alpha f}(\tilde{y})\|_2^2 \geq \|\tilde{x} - \tilde{y}\|_2^2$, as illustrated in Figure 2.
- When $\tilde{x} = 0$ and $d > 0$, $\mathrm{prox}_{\alpha f}$ is multivalued in 0 as all the points on the circle of radius ρ^+ are solution of (12). As a consequence, f is not *prox-regular* at $\{0\}$ and its proximity operator is not defined for this point.

For practical reasons, we define $\mathrm{prox}_{\alpha f}$ everywhere by assuming that $\angle(0) = 0$. Thus, the proximity operator of f is

$$\mathrm{prox}_{\alpha\, f}(\tilde{x}) = \begin{cases} \rho^+ & \text{if } \tilde{x} = 0\,, \\ \rho^+ \exp(\jmath\, \tilde{\phi}) & \text{otherwise}\,, \end{cases} \tag{15}$$

Let us notice that the modified projection $P'(\tilde{x}, d)$ defined by Equation 8 lies also on the line between $\tilde{x}$ and its projection on the circle $|x|^2 = d$. Its position on this line depend on the value of the relaxation parameter β. We can thus reinterpret this modified projection as a heuristic approximation of the proximity operator.

C. Gaussian Likelihood

For additive Gaussian noise at a given pixel with variance $\sigma = \mathrm{Var}\{d\}$, the function f writes

$$f(x) = w(|x|^2 - d)^2\,, \tag{16}$$

where $w = 1/\sigma^2$ is the inverse variance of the noise at the considered pixel. In this case, (13) becomes:

$$\rho^+ = \underset{\rho\geq 0}{\arg\min} \left(\alpha\, w\, (\rho^2 - d)^2 + \frac{1}{2}\, (\rho - \tilde{\rho})^2 \right). \tag{17}$$

The solution is then one of the roots of the polynomial q_G defined as

$$\begin{aligned} q_G(\rho) &= \frac{\mathrm{d}}{\mathrm{d}\rho} \left(\alpha\, w\, (\rho^2 - d)^2 + \frac{1}{2}\, (\rho - \tilde{\rho})^2 \right) \\ &= 4\, \alpha\, w\, \rho^3 + \rho\, (1 - 4\, \alpha\, wd) - \tilde{\rho}\,. \end{aligned} \tag{18}$$

As there is no second coefficient in this cubic polynomial, the sum of its roots is zero whereas their product is strictly positive since $\tilde{\rho}/(4\, \alpha\, w) > 0$. Thus, q_G has always only one positive root ρ^+. As stated in the previous section, this root must lie between $\sqrt{d}$ and $\tilde{\rho}$. It is computed using Cardano's method.

D. Poisson Likelihood

In the photon-counting case, the noise follows a Poisson distribution and the function f writes

$$f(x) = |x|^2 - d\, \log\left(|x|^2 + b\right), \tag{19}$$

where b is the expectation of some spurious independent Poisson process that accounts for background emission and detector dark current at the considered pixel. Given this noise distribution, the solution of (13) is given by the largest real root of the cubic polynomial $q_P(\rho) = \frac{\mathrm{d}f(\rho)}{\mathrm{d}\rho}$, with

$$\begin{aligned} q_P(\rho) &= \frac{\mathrm{d}}{\mathrm{d}\rho} \left(\alpha\, f(\rho) + \frac{1}{2}\, (\rho - \tilde{\rho})^2 \right) \\ &= (2\, \alpha + 1)\, \rho^3 - \tilde{\rho}\, \rho^2 + ((2\, \alpha + 1)\, b - 2\, \alpha\, d)\, \rho - b\, \tilde{\rho}\,. \end{aligned} \tag{20}$$

As in the case of (18), this root is computed using Cardano's method. When no background emission is present ($b = 0$), this polynomial reduces to a quadratic equation whose largest root always exists and is given by

$$\rho^+ = \frac{\tilde{\rho} + \sqrt{8\, d\, \alpha\, (1 + 2\, \alpha) + \tilde{\rho}^2}}{2 + 4\, \alpha}\,. \tag{21}$$

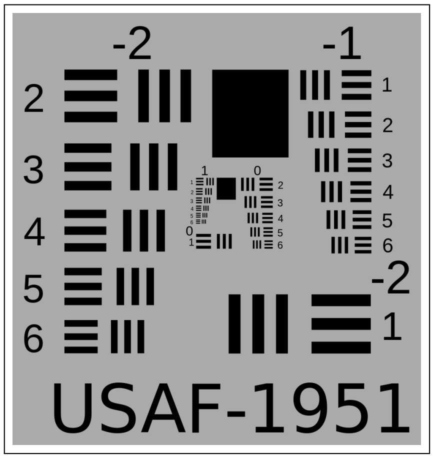

Fig. 3. USAF-1951 test image used.

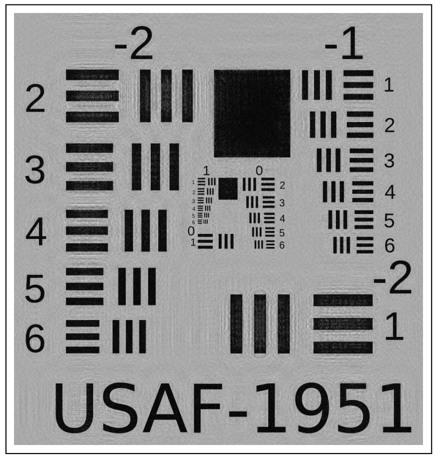

Fig. 4. Intensity of the estimated wave using the classical projection (noiseless case) back-propagated from z_A to z_0. $\mathrm{SNR}(\boldsymbol{x}^+) = 15.12\,\mathrm{dB}$.

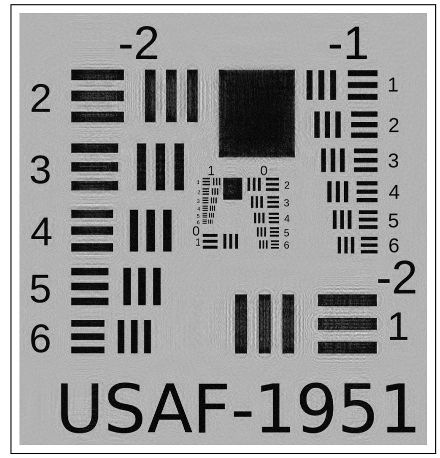

Fig. 5. Intensity of the estimated wave using the proposed proximity operator (noiseless case) back-propagated from z_A to z_0. $\mathrm{SNR}(\boldsymbol{x}^+) = 15.68\,\mathrm{dB}$.

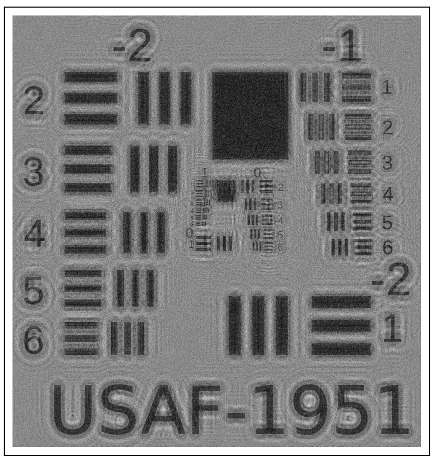

Fig. 6. Intensity of the estimated wave using the classical projection (noise standard deviation $\sigma = 0.3$) back-propagated from z_A to z_0. $\mathrm{SNR}(\boldsymbol{x}^+) = 6.37\,\mathrm{dB}$.

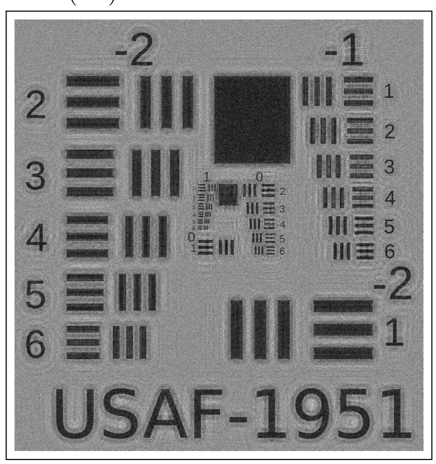

Fig. 7. Intensity of the estimated wave using the proposed proximity operator (noise standard deviation $\sigma = 0.3$) back-propagated from z_A to z_0. $\mathrm{SNR}(\boldsymbol{x}^+) = 7.60\,\mathrm{dB}$.

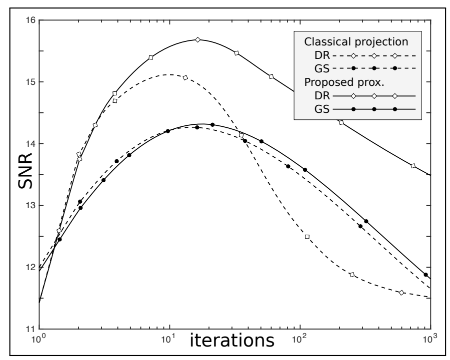

Fig. 8. Comparison of DR and GS performance without noise using the classical projection or the proposed operator

4. PROXIMITY OPERATOR FOR A SUM OF INTENSITY MEASUREMENTS

In this section, we extend the presented proximity operators to the case where N complex amplitudes sum up incoherently on a pixels. This corresponds to the multispectral case or when interference fringes exhibit high frequencies that are not sufficiently sampled by the detector. In this case, an appropriate forward model is

$$d_k = \|\boldsymbol{y}_k\|_2^2 + n_k\,, \tag{22}$$

where $\boldsymbol{y}_k \in \mathbb{C}^N$ is a vector containing the N complex amplitudes arriving on the pixels k. In the undersampled-fringes case, this vector writes $\boldsymbol{y}_k = (x_{N(k-1)+1}, \ldots, x_{Nk})$, where the factor N is chosen such that the adequately sampled complex amplitude $\boldsymbol{x} \in \mathbb{C}^{NK}$ fulfills the Nyquist criterion. With this forward model, the likelihood function writes $\ell_k(\|\boldsymbol{y}_k\|^2\,;\,d_k)$. By setting $\boldsymbol{y}_k = \eta\,\boldsymbol{u}$, with $\eta \geq 0$ and $\|\boldsymbol{u}\|_2 = 1$, we can define $\alpha\,f(\eta) = \ell_k(\eta^2\,;\,d_k)$.

Then (12) becomes

$$\mathrm{prox}_{\alpha f}(\widetilde{\boldsymbol{y}}) = \operatorname*{arg\,min}_{\eta \geq 0\,,\;\|\boldsymbol{u}\|_2 = 1} \left(\alpha\,f(\eta) + \frac{1}{2}\,\|\eta\,\boldsymbol{u} - \widetilde{\boldsymbol{y}}\|_2^2 \right)\,. \tag{23}$$

By assuming that $\|\widetilde{\boldsymbol{y}}\|_2 \neq 0$ and $\eta > 0$, we find that

$$\boldsymbol{u}^+(\eta) = \operatorname*{arg\,min}_{\boldsymbol{u},\|\boldsymbol{u}\|=1} \|\eta\,\boldsymbol{u} - \widetilde{\boldsymbol{y}}\|_2^2 = \frac{\widetilde{\boldsymbol{y}}}{\|\widetilde{\boldsymbol{y}}\|_2}\,. \tag{24}$$

Thus the solution is

$$\boldsymbol{y}^+ = \eta^+\,\frac{\widetilde{\boldsymbol{y}}}{\|\widetilde{\boldsymbol{y}}\|_2}\,, \tag{25}$$

where η^+ is given by

$$\eta^+ = \operatorname*{arg\,min}_{\eta} \left(\min_{\boldsymbol{u},\|\boldsymbol{u}\|=1} \left(\alpha\,f(\eta) + \frac{1}{2}\,\|\eta\,\boldsymbol{u} - \widetilde{\boldsymbol{y}}\|_2^2 \right) \right)\,, \tag{26}$$

$$= \operatorname*{arg\,min}_{\eta > 0} \left(\alpha\,f(\eta) + \frac{1}{2}\,(\eta - \|\widetilde{\boldsymbol{y}}\|_2)^2 \right)\,, \tag{27}$$

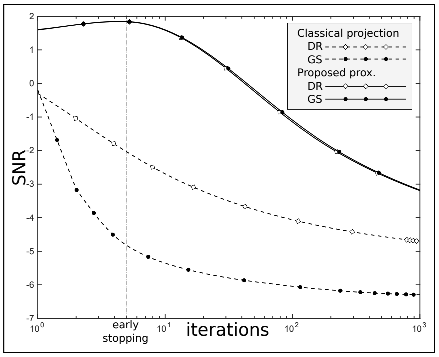

Fig. 9. Comparison of DR and GS performance in noisy conditions ($\sigma = 1$) using the classical projection or the proposed operator.

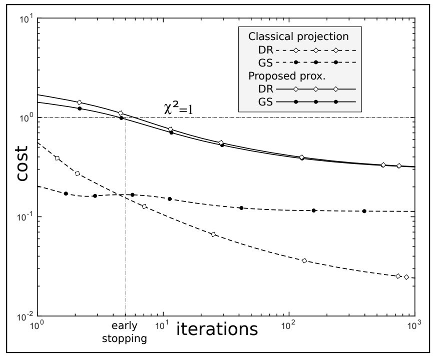

Fig. 10. Evolution of the cost function for both algorithms and both projectors.

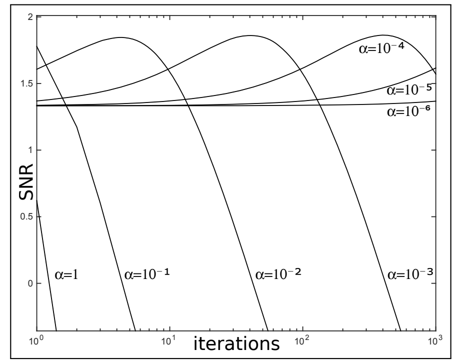

Fig. 11. Performance of DR ($\sigma = 1$) with the proposed proximity operator $\mathrm{prox}_{\alpha f}$ for different values of α.

since

$$\min_{\boldsymbol{u}, \|\boldsymbol{u}\|=1} \|\eta\, \boldsymbol{u} - \widetilde{\boldsymbol{y}}\|_2^2 = (|\eta| - \|\widetilde{\boldsymbol{y}}\|_2)^2 \,. \tag{28}$$

Solving (27) is equivalent to solving (13) with $\widetilde{\rho} = \|\widetilde{\boldsymbol{y}}\|_2$. In the case where $\|\widetilde{\boldsymbol{y}}\|_2 = 0$ and $d > 0$, f is not prox-regular and (23) has an infinite number of solutions. As in Section B, we assume in practice that $\mathrm{prox}_{\alpha f}(\widetilde{\boldsymbol{y}}) = \eta^+$ when $\|\widetilde{\boldsymbol{y}}\|_2 = 0$. To sum up, the proximity operator for undersampled measurements is:

$$\mathrm{prox}_{\alpha f}(\widetilde{\boldsymbol{y}}) = \begin{cases} \eta^+ \,, & \text{if } \|\widetilde{\boldsymbol{y}}\|_2 = 0 \\ \eta^+ \frac{\widetilde{\boldsymbol{y}}}{\|\widetilde{\boldsymbol{y}}\|_2} \,, & \text{otherwise} \,. \end{cases} \tag{29}$$

This proximity operator for undersampled intensity measurements can be computed for any function f that has a proximity operator in closed form such the Gaussian or Poisson likelihood described in the previous sections.

5. NUMERICAL EXPERIMENTS

To study the performance of the proposed proximity operators, we simulated one of the simplest setup of phase retrieval. Under a Fresnel approximation, we simulated numerically a wave diffracted by a planar real object (here a $K = 1024 \times 984$ pixels image of the USAF resolution test chart shown Figure 3) placed at $z_0 = 0$. The diffracted wave at z_A is the reference complex amplitude $\boldsymbol{r}$ that will be estimated throughout the experiments. We computed the noisy intensities $\boldsymbol{d}_A = |\boldsymbol{r}| + \boldsymbol{n}_A$ and $\boldsymbol{d}_B = |\mathbf{H}\,\boldsymbol{r}| + \boldsymbol{n}_B$ at depth z_A and z_B, where $\mathbf{H}$ is the propagation operator from z_A to z_B and $\boldsymbol{n}_A$ and $\boldsymbol{n}_B$ are noise vectors with identical statistics given by the experimental conditions. The setup parameters are: $\lambda = 633\,\mathrm{nm}$, pixel size: $= 5.3\,\mu\mathrm{m}$, $z_A = 1\,\mathrm{cm}$, and $z_B = 2\,\mathrm{cm}$.

For each experiment, we built the functions $f_{A,k}(x) = \ell_k(|x|^2\,;\, d_{A,k})$ and $f_{B,k}(x) = \ell_k(|x|^2\,;\, d_{B,k})$ according to the considered noise model. We then compared the performance of the proposed proximity operator $\mathrm{prox}_{\alpha f}$ to that of the classical projection defined by (6) by estimating the complex amplitude of the wave $\boldsymbol{x}^+$ at z_A. To keep the problem as simple as possible, we only used the knowledge of measured intensities without additional prior (neither regularization, nor use of the fact that the image is non-negative at z_0).

In all experiments, the quality of the recovered complex amplitude $\boldsymbol{x}$ in plane z_A is assessed by the mean of the reconstruction signal to noise ratio:

$$\mathrm{SNR}(\boldsymbol{x}) = 10\,\log_{10} \frac{\|\boldsymbol{r}\|_2^2}{\|\boldsymbol{r} - \boldsymbol{x}\|_2^2} \,. \tag{30}$$

As the initial wave is real in the plane $z_0 = 0$, back-propagating the estimated wave from z_A to z_0 is used as a visual assessment of the reconstruction quality as shown Figures 4 to 7. Let us remind that as the phase retrieval problem is not convex, the solution depends on the initialization. We chose the initialization $\boldsymbol{x}^{(0)} = \sqrt{\boldsymbol{d}_A}$ for every experiments and a different initialization may lead to a different recovered complex amplitude with a different SNR.

A. Alternating Projection or Douglas Rachford?

Algorithm 2. Douglas-Rachford algorithm

procedure DR(f_A, f_B)
  $\boldsymbol{y}^{(0)} = \sqrt{\boldsymbol{d}_A}$ and $\lambda \in]0, 2[$ ▷ init. ($\lambda = 1$ for all results)
  for $n = 1, \ldots,$ maxiter **do**
    $\boldsymbol{x}^{(n)} = \mathrm{prox}_{\alpha f_A}(\boldsymbol{y}^{(n-1)})$
    $\boldsymbol{r}^{(n)} = 2\,\boldsymbol{x}^{(n)} - \boldsymbol{y}^{(n-1)}$
    $\boldsymbol{y}^{(n)} = \boldsymbol{y}^{(n-1)} + \lambda \left(\mathbf{H}^\top \mathrm{prox}_{\alpha f_B}\left(\mathbf{H}\,\boldsymbol{r}^{(n)}\right) - \boldsymbol{x}^{(n)}\right)$
  return $\boldsymbol{x}^{(\mathrm{maxiter})}$ ▷ Complex amplitude in the z_A plane

The use of the proposed operator in Algorithm 1 instead of the classical projection P_A and P_B amounts to solving

$$\begin{aligned} \boldsymbol{x}^+ \;\in\; & \operatorname*{arg\,min}_{\boldsymbol{x}\in\mathbb{C}^K} \left(\sum_{k=1}^{N} f_{A,k}(x_k) \right. \\ & \left. + \inf_{\boldsymbol{y}\in\mathbb{C}^K} \left(\sum_{k'=1}^{K} f_{B,k}(y_k) + \frac{1}{2}\,\|\mathbf{H}\,\boldsymbol{x} - \boldsymbol{y}\|_2^2 \right) \right) \,, \end{aligned} \tag{31}$$

which is a relaxed version of (10). Alternatively, (10) can be solved using the Douglas-Rachford (DR) algorithm described in Alg. 2 thanks to the following property on the proximity operator of $g(\boldsymbol{x}) = f(\mathbf{H} \cdot \boldsymbol{x})$ [26]:

$$\mathbf{H} \cdot \mathbf{H}^\top = \mathbf{Id} \Longrightarrow \mathrm{prox}_{\alpha g}(\boldsymbol{x}) = \mathbf{H}^\top \cdot \mathrm{prox}_{\alpha f}(\mathbf{H} \cdot \boldsymbol{x}) \,, \tag{32}$$

where $\mathbf{Id}$ is the identity matrix.

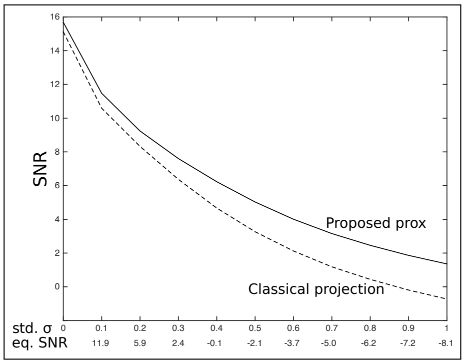

Fig. 12. Comparison of both projectors using the DR algorithm as a function of noise (noise level given in standard deviation and SNR).

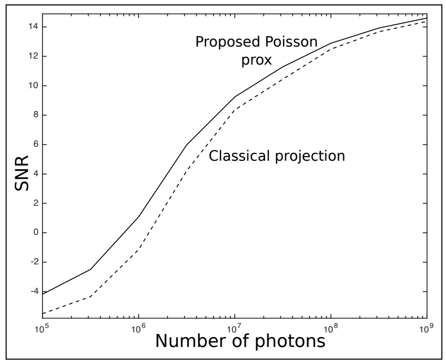

Fig. 13. Comparison of classical projection and Poisson proximity as a function of the number of photons (10^5 photons ≈ 1 photon per pixel, on average).

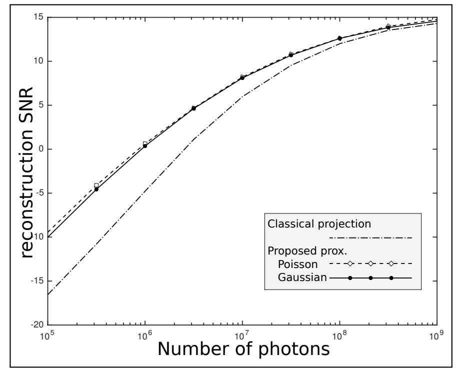

Fig. 14. $\mathrm{SNR}(\boldsymbol{x}^+)$ as function of the number of photons in presence of a $3\,e^-$ dark current.

For the Gaussian likelihood as for the Poisson likelihood, f is not convex. The convergence of both algorithms cannot be proved even if there exist some convergence results in the related case of the estimation of the intersection of a circle and a line [35]. The solution may therefore depend on the starting point. In all the presented experiments, we begin with the starting amplitude in z_A plane $x_A^{(0)} = \sqrt{d_A}$.

With the classical projection, DR is more efficient than GS as can be seen in Figure 8 and Figure 9, either with or without noise. In the presence of noise and using the proposed proximity operator, the performances of both algorithms are similar; they become indistinguishable as the amounts of noise level increases.

B. Tuning the Parameters

With the proposed proximity operator, two parameters have to be tuned: the number of iterations and the parameter α. All tests with the DR algorithm were done with $\lambda = 1$.

Phase retrieval is an ill-posed problem. The number of unknowns ($2\,K$) is equal to the number of measurements, meaning that such *maximum-likelihood* algorithms are subject to noise amplification. Hence, $\mathrm{SNR}(\boldsymbol{x})$ began to worsen after some iteration, while the cost was still decreasing, as can be seen Figure 9 and Figure 10. The correct prescription of the number of iterations is essential to stop the algorithm at the precise moment when the wavefront gives the best SNR. This is classically done in phase retrieval and acts as a regularization [36]. To set the maximum number of iterations, we apply the Morozov principle; the algorithm only proceeds as long as:

$$\chi^2 = \frac{1}{2\,K}\left(\sum_{k=1}^{K} f_{A,k}\left(x_k\right) + \sum_{k'=1}^{K} f_{B,k'}\left([\mathbf{H}\,\boldsymbol{x}]_{k'}\right)\right) < 1\,. \qquad \textbf{(33)}$$

In our experiments, this criterion seems to stop the algorithm close to the optimum, as can be seen in Figure 9 and Figure 10.

From Figure 11, it can be seen that the parameter α has a strong effect on the speed of convergence but has little influence on the quality. However, if α is too large (*e.g.* $\alpha = 1$ in Figure 11), the steps are too large and the criterion χ^2 is well below 1 even after the first iteration. As consequence, α is set such that $\chi^2 > 1$ for the first few iterations.

Such an automatic tuning works only for the Gaussian likelihood. In the absence of noise, for the Poisson likelihood and the classical projection, we select the number of iterations and α that maximize $\mathrm{SNR}(\boldsymbol{x}^+)$.

C. Gaussian Noise

We first compare the classical projection with the proximity operator derived from the Gaussian likelihood. In the noiseless case, the proximity operator improves $\mathrm{SNR}(\boldsymbol{x}^+)$ by about 0.5 dB. However, the visual differences between both reconstructions back-projected in the z_0 plane are barely noticeable as shown on Figures 4 and 5.

For the noisy scenario, the reconstruction error as a function of the standard deviation of the noise is shown in Figure 12. We observe that the use of the proximity operator always improves $\mathrm{SNR}(\boldsymbol{x}^+)$ by at least 0.5 dB compared to the classical projection. When the noise is $\sigma = 0.3$ or higher (*i.e.*, the SNR of the measurements is lower than 2.4 dB), the classical projection fails to properly estimate any phase. As consequence, the twin image appears much more clearly in the back-propagated field to z_0 in the classical projection case than with the proposed proximity operator, as can be seen in Figure 6 and Figure 7.

D. Photon Counting

To test the proximity operator derived for the Poisson likelihood we performed simulations while varying the illumination and without any background emission ($b_k = 0$), in which case the proximity operator is given by (21). We compared its performance to that of the classical projection for an illumination varying from 10^5 to 10^9 photons in each plane. Compared to the classical projection, the proposed proximity operator always improves $\mathrm{SNR}(\boldsymbol{x}^+)$, as can be seen in Figure 13. The performance gap with to classical projection becomes smaller as the number of photons increases.

E. Low-Light Conditions

In low light, most detection devices are plagued by dark current, which can be modeled by an additive background emission $b_k > 0$. For illuminations from 10^5 to 10^9 photons, we simulated the measured intensity d_k at pixel k following a Poisson distribution $\mathcal{P}$, so that

$$d_k = \mathcal{P}\left(|x_k|^2 + b\right)\,, \qquad \textbf{(34)}$$

where the dark current was set to $b = 3\,e^-$ per pixel. The reconstruction SNR as a function of illumination is shown on Figure 14 for the classical projection, the Poisson-likelihood proximity operator and the Gaussian-likelihood proximity operator assuming

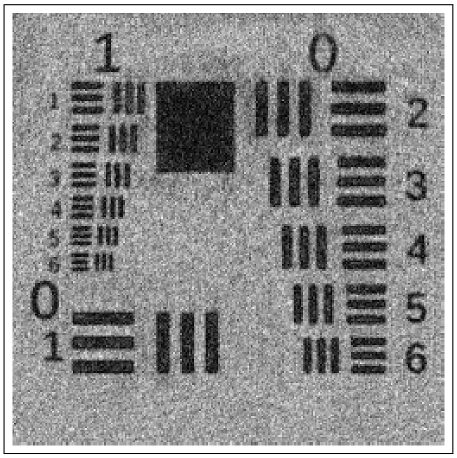

Fig. 15. Central 250 × 250 pixels of the test chart recovered from 8 planes measurements using the proximity operator for sum-of-intensity measurements and back-propagated to $z = 0$.

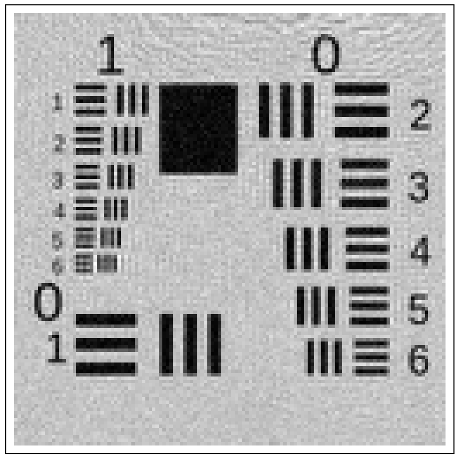

Fig. 16. Central 125 × 125 pixels of the test chart recovered from 8 planes measurements using the proximity operator presented in Section C and back-propagated to $z = 0$.

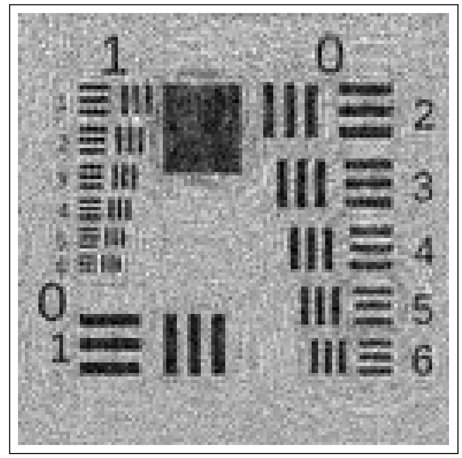

Fig. 17. Central 125 × 125 pixels of the test chart recovered from the 2 planes measurements using the proximity operator presented in Section C and back-propagated to $z = 0$.

a signal-dependent Gaussian noise with mean $b = 3$ and an inverse variance w_k at pixel k estimated as

$$w_k = 1/\max(d_k, b)\,. \tag{35}$$

In Figure 14, we see that the two proximity operators have a very similar performance and perform better than the classical projection. Interestingly, both proximity operators. This means that, even with a quite low dark current (here $b = 3$), the approximation of a Poisson noise with the non-stationary Gaussian noise given in (35) is good.

F. Undersampled Fringes: Trading SNR for Resolution.

We tested the sum-of-intensity proximity operator derived in section 4 in the case where the fringes are not sufficiently sampled by the detector. Given the adequately sampled complex amplitude $\boldsymbol{g}_p \in \mathbb{C}^{K_1 \times K_2}$ in the detector plane z_p, we simulated (2×2) subsampled intensity measurements $\boldsymbol{d}_p \in \mathbb{R}^{M_1 \times M_2}$ with $K = 2\,M$ using the direct model

$$\begin{aligned} g_p &= \mathbf{H}_p \cdot \boldsymbol{r}\,, && (36)\\ d_{p,m_1,m_2} &= \left|g_{p,2\,m_1,2m_2}\right|^2 + \left|g_{p,2\,m_1+1,2m_2}\right|^2 \\ &\quad + \left|g_{p,2\,m_1,2\,m_2+1}\right|^2 + \left|g_{p,2\,m_1+1,2\,m_2+1}\right|^2 + \boldsymbol{n}_p && (37) \end{aligned}$$

where $\mathbf{H}_p$ is the propagation operator from the plane z_1 to z_p. As in the previous experiments, we estimated the complex amplitude $\boldsymbol{x}^+$ in the plane z_1.

The strategy without regularization is only viable when there are sufficiently many measurements $(P \times M_1 \times M_2)$ as compared to the number of unknowns $(2 \times K_1 \times K_2 = 8 \times M_1 \times M_2)$. To increase the number of measurements, we modified the proposed setup and estimated $\boldsymbol{x}^+$ in the plane z_1 from $P = 8$ measurements.

The maximum-likelihood solution in this case is given by

$$\boldsymbol{x}^+ \in \underset{\boldsymbol{x} \in \mathbb{C}^K}{\arg\min} \sum_{p=1}^{8} \sum_{k=1}^{K} f_{p,k}\left(\left[\mathbf{H}_p \cdot \boldsymbol{x}\right]_k\right)\,. \tag{38}$$

It is solved by means of the PPXA algorithm [37], which is a generalization of the Douglas-Rachford algorithm that minimizes the sum of more than two functions.

We simulated intensity measurements for eight planes taken at $z_1 = 1\,\text{cm}$, $z_2 = 1.5\,\text{cm}$, $z_3 = 2\,\text{cm}$, $z_4 = 2.5\,\text{cm}$, $z_5 = 3\,\text{cm}$, $z_6 = 3.5\,\text{cm}$, $z_7 = 4\,\text{cm}$ and $z_8 = 4.5\,\text{cm}$. These measurements were corrupted with additive Gaussian noise of variance $\sigma = 0.5$ (coresponding to $\text{SNR}() = -2.1\,\text{dB}$).

We have estimated the 1024×968 pixels complex amplitude $\boldsymbol{x}^+$ in the plane z_1 from these eight 512×484 pixels intensity measurements using the proposed proximity operator for sum of intensities with f derived for the Gaussian likelihood (16). A zoom on the central part of the wave back-propagated to z_0 is presented in Figure 15. It illustrates the effectiveness of the proposed proximity operator to recover fine details and increase the resolution. This can be compared with two reconstructions without superresolution using the same PPXA algorithm but with the proximity operator derived in Section C. One, shown on Figure 16, was done with the same measurements ($8 \times M_1 \times M_2$ measurements for $2 \times M_1 \times M_2$ unknowns). The other, shown on Figure 17, is using only the measurements in the two planes z_1 and z_2 to get the same number measurements than unknowns $(2 \times M_1 \times M_2)$. Compared to these non-superresolved reconstructions, the resolution improvement is obvious. However, this improvement is acquired at the cost of a moderate increase in noise compared to the reconstruction shown on the Figure 16. Indeed, the non-superresolved reconstruction appears less noisy as the ratio of the number of unknowns over the number of measurements is more favorable. This reconstruction noise is similar to in the non-superresolved reconstruction using only two planes to get the same number measurements than unknowns shown on the Figure 17.

6. CONCLUSION

We considered the problem of the phase retrieval from noisy intensity measurements. From the maximum-likelihood formulation, we derived proximal operators for intensity measurements corrupted with Gaussian noise or Poisson noise. We further expanded these proximity operators for cases where fringes are not properly sampled. When plugged into the Gerchberg-Saxton algorithm in place of the classical projection, it showed superior results. As it can be plugged into any projection-based algorithm, it can provide an improvement of the performance for many phase-retrieval algorithms without changing the core of the optimization procedure.

ACKNOWLEDGEMENTS

This work is supported by the Sinergia project "Euclid: precision cosmology in the dark sector" from the Swiss National Science Foundation and by the French ANR POLCA project (Processing of pOLychromatic interferometriC data for Astrophysics, ANR-10-BLAN-0511).

REFERENCES

1. A. Walther, "The question of phase retrieval in optics," Journal of Modern Optics **10**, 41–49 (1963).
2. R. W. Harrison, "Phase problem in crystallography," JOSA A **10**, 1046–1055 (1993).
3. D. Misell, "A method for the solution of the phase problem in electron microscopy," Journal of Physics D: Applied Physics **6**, L6 (1973).
4. C. Fienup and J. Dainty, "Phase retrieval and image reconstruction for astronomy," Image Recovery: Theory and Application pp. 231–275 (1987).
5. R. W. Gerchberg and W. O. Saxton, "A practical algorithm for the determination of phase from image and diffraction plane pictures," Optik **35**, 237 (1972).
6. 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," Signal Processing Magazine, IEEE **32**, 87–109 (2015).
7. J. Fienup, "Iterative method applied to image reconstruction and to computer-generated holograms," Optical Engineering **19**, 193297–193297 (1980).
8. J. R. Fienup, "Phase retrieval algorithms: A comparison," Applied Optics **21**, 2758–2769 (1982).
9. V. Elser, "Solution of the crystallographic phase problem by iterated projections," Acta Crystallographica Section A: Foundations of Crystallography **59**, 201–209 (2003).
10. H. H. Bauschke, P. L. Combettes, and D. R. Luke, "Hybrid projection–reflection method for phase retrieval," JOSA A **20**, 1025–1034 (2003).
11. D. R. Luke, "Relaxed averaged alternating reflections for diffraction imaging," Inverse Problems **21**, 37 (2005).
12. E. J. Candès, T. Strohmer, and V. Voroninski, "Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming," Communications on Pure and Applied Mathematics **66**, 1241–1274 (2013).
13. F. Fogel, I. Waldspurger, and A. d'Aspremont, "Phase retrieval for imaging problems," arXiv preprint arXiv:1304.7735 (2013).
14. Y. Shechtman, A. Beck, and Y. C. Eldar, "Gespar: Efficient phase retrieval of sparse signals," IEEE Transactions on Signal Processing, **62**, 928–938 (2014).
15. T. T. Cai, X. Li, and Z. Ma, "Optimal rates of convergence for noisy sparse phase retrieval via thresholded wirtinger flow," arXiv preprint arXiv:1506.03382 (2015).
16. A. M. Tillmann, Y. C. Eldar, and J. Mairal, "Dictionary learning from phaseless measurements," in "2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)," (2016), pp. 4702–4706.
17. E. J. Candes, X. Li, and M. Soltanolkotabi, "Phase retrieval via wirtinger flow: Theory and algorithms," IEEE Transactions on Information Theory **61**, 1985–2007 (2015).
18. Y. Chen and E. Candes, "Solving random quadratic systems of equations is nearly as easy as solving linear systems," in "Advances in Neural Information Processing Systems," (2015), pp. 739–747.
19. A. Drémeau and F. Krzakala, "Phase recovery from a bayesian point of view: The variational approach," in "IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)," (IEEE, 2015), pp. 3661–3665.
20. G. Williams, M. Pfeifer, I. Vartanyants, and I. Robinson, "Effectiveness of iterative algorithms in recovering phase in the presence of noise," Acta Crystallographica Section A: Foundations of Crystallography **63**, 36–42 (2007).
21. T. Latychevskaia, J.-N. Longchamp, and H.-W. Fink, "Novel Fourier-domain constraint for fast phase retrieval in coherent diffraction imaging," Optics Express **19**, 19330–19339 (2011).
22. R. A. Dilanian, G. J. Williams, L. W. Whitehead, D. J. Vine, A. G. Peele, E. Balaur, I. McNulty, H. M. Quiney, and K. A. Nugent, "Coherent diffractive imaging: A new statistically regularized amplitude constraint," New Journal of Physics **12**, 093042 (2010).
23. A. V. Martin, F. Wang, N. D. Loh, T. Ekeberg, F. R. N. C. Maia, M. Hantke, G. van der Schot, C. Y. Hampton, R. G. Sierra, A. Aquila, S. Bajt, M. Barthelmess, C. Bostedt, J. D. Bozek, N. Coppola, S. W. Epp, B. Erk, H. Fleckenstein, L. Foucar, M. Frank, H. Graafsma, L. Gumprecht, A. Hartmann, R. Hartmann, G. Hauser, H. Hirsemann, P. Holl, S. Kassemeyer, N. Kimmel, M. Liang, L. Lomb, S. Marchesini, K. Nass, E. Pedersoli, C. Reich, D. Rolles, B. Rudek, A. Rudenko, J. Schulz, R. L. Shoeman, H. Soltau, D. Starodub, J. Steinbrener, F. Stellato, L. Strüder, J. Ullrich, G. Weidenspointner, T. A. White, C. B. Wunderer, A. Barty, I. Schlichting, M. J. Bogan, and H. N. Chapman, "Noise-robust coherent diffractive imaging with a single diffraction pattern," Optics Express **20**, 16650–16661 (2012).
24. A. Schutz, A. Ferrari, D. Mary, F. Soulez, É. Thiébaut, and M. Vannier, "Painter: a spatiospectral image reconstruction algorithm for optical interferometry," JOSA A **31**, 2334–2345 (2014).
25. D. Weller, A. Pnueli, G. Divon, O. Radzyner, Y. Eldar, and J. Fessler, "Undersampled phase retrieval with outliers," IEEE Transactions on Computational Imaging, **1**, 247–258 (2015).
26. P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in "Fixed-Point Algorithms for Inverse Problems in Science and Engineering," (Springer, 2011), pp. 185–212.
27. D. Noll and A. Rondepierre, "On local convergence of the method of alternating projections," Foundations of Computational Mathematics pp. 1–31 (2015).
28. R. G. Lyon, J. E. Dorband, and J. M. Hollis, "Hubble space telescope faint object camera calculated point-spread functions," Appl. Opt. **36**, 1752–1765 (1997).
29. D. R. Luke, J. V. Burke, and R. G. Lyon, "Optical wavefront reconstruction: Theory and numerical methods," SIAM review **44**, 169–224 (2002).
30. A. Levi and H. Stark, "Image restoration by the method of generalized projections with application to restoration from magnitude," JOSA A **1**, 932–943 (1984).
31. H. H. Bauschke, P. L. Combettes, and D. R. Luke, "Phase retrieval, error reduction algorithm, and Fienup variants: A view from convex optimization," Journal of Optical Society of America A **19**, 1334–1345 (2002).
32. A. Repetti, E. Chouzenoux, and J.-C. Pesquet, "A nonconvex regularized approach for phase retrieval," in "IEEE International Conference on Image Processing (ICIP)," (IEEE, 2014), pp. 1753–1757.
33. J.-J. Moreau, "Proximité et dualité dans un espace hilbertien," Bulletin de la Société mathématique de France **93**, 273–299 (1965).
34. W. Hare and C. Sagastizábal, "Computing proximal points of nonconvex functions," Mathematical Programming **116**, 221–258 (2009).
35. R. Hesse and D. R. Luke, "Nonconvex notions of regularity and convergence of fundamental algorithms for feasibility problems," SIAM Journal on Optimization **23**, 2397–2419 (2013).
36. D. R. Luke, "Local linear convergence of approximate projections onto regularized sets," Nonlinear Analysis: Theory, Methods & Applications **75**, 1531–1546 (2012).
37. N. Pustelnik, C. Chaux, and J.-C. Pesquet, "Parallel proximal algorithm for image restoration using hybrid regularization," IEEE Transactions on Image Processing **20**, 2450–2462 (2011).