Proximity Operators for Phase Retrieval
Ferréol Soulez, Éric Thiébaut, Antony Schutz, André Ferrari, Frédéric Courbin, Michael Unser

To cite this version:

HAL Id: insu-01632688
https://hal-insu.archives-ouvertes.fr/insu-01632688
Submitted on 13 Nov 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Proximity Operators for Phase Retrieval

FERREOL SOULEZ, ÉRICK THIEBAUT, ANTONY SCHUTZ, ANDRE FERRARI, FREDERIC COURBIN, AND MICHAEL UNSER

1 Biomedical Imaging Group, École polytechnique fédérale 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, Université 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

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.

1. INTRODUCTION

The classical phase-retrieval problem is to reconstruct a complex-valued signal $x$ from measurements of its squared modulus $|x|^2$. This problem arises in many applications (e.g., crystallography [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 $x = (x_1, \ldots, x_K)$, where $K$ is the number of pixels. In the phase retrieval problem, the forward model that links the complex amplitude $x \in \mathbb{C}^K$ to the measured image intensities $d \in \mathbb{R}_+^K$ is

$$d_k = |x_k|^2 + n_k, \quad (1)$$

where $n$ 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 $x$ that minimizes a cost function

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

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 $x$ (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 \( d_A \) and \( d_B \) measured at depth \( z_A \) and \( z_B \), respectively. It solves the feasibility problem

\[
\text{find } x \in C_A \cap \{ x : H x \in C_B \},
\]

where \( H \) 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 \( d_i \), i.e., \( C_i = \{ x \in \mathbb{C}^n, |x|^2 = d_i \} \). The propagation operator is classically either the Fresnel operator (under a Fresnel approximation) or the Fourier operator (under a Fraunhofer approximation). This can be reformulated as the minimization problem

\[
x^+ \in \arg \min_{x \in \mathbb{C}^n} \{ t_{C_A}(x) + t_{C_B}(H x) \},
\]

where \( t_C \) is the indicator function of the set \( C \) defined as

\[
t_C(x) = \begin{cases} 
0, & \text{if } x \in C \\
\infty, & \text{otherwise}
\end{cases}
\]

Observe that, when \( d_i \neq 0 \), \( C_i \) is generally not a convex set. Therefore, only local convergence can be established [27].

Algorithm 1. Gerchberg-Saxton algorithm

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

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

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

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

\[
\min_{y \in \mathbb{C}^n} \{ t_C(y) + \frac{1}{2} \| x - y \|^2 \}.
\]

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

\[
P'(x_k | d_k) = (1 - \beta) x_k + \beta P(x_k | d_k),
\]

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 log-likelihood of the noise distribution at pixel \( k \) (up to the constant \( \text{cst} \)) is:

\[
\ell_k(n_k) = -\ln \Pr(n_k | x_k) + \text{cst}.
\]

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

\[
x^+ \in \arg \min_{x \in \mathbb{C}^n} \left( \sum_{k=1}^{K} \ell_k(|x_k|^2 - d_k) + \sum_{k'=1}^{K} \ell_k\left(|H x|_{k'}^2 - d_k\right) \right).
\]

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 \( \ell \) [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 \( t_C \) in (7) is replaced by an arbitrary lower semi-continuous convex function \( g : \mathbb{C}^n \to \mathbb{R} \) so that

\[
\text{prox}_g(x) \overset{\text{def}}{=} \arg \min_{y \in \mathbb{C}^n} \{ g(y) + \frac{1}{2} \| x - y \|^2 \}.
\]

The concept of proximal mapping has also been extended to non-convex functions that fulfill three conditions: (i) lower semi-continuity; (ii) \( \text{prox-boundedness} \); and (iii) \( \text{prox-regularity} \) (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 \( a f(x) = \ell_k(d_k - |x|^2) \) with \( a > 0 \) a tuning factor. The function \( f \) has the following properties: (i) continuity, provided that \( \ell_k \) is also continuous (that is true for most noise statistics used in practice); (ii) non convexity (e.g., if \( x_1 = \sqrt{a} \) is a minimum of \( f \), then \( x_2 = -\sqrt{a} \) is, but not necessarily \( (x_1 + x_2)/2 \)); (iii) \( \text{prox-boundedness} \) as \( f \) is positive (and
The solution \( x^+ (\alpha) \) is tangent. However, as described further, \( f \) is not \textit{prox-regular} in \( x = 0 \).

The proximity operator of \( a \, f \) is given by

\[
\text{prox}_{a \, f} (\bar{x}) = \arg \min_{x \in \mathbb{C}} \left\{ a \, f (x) + \frac{1}{2} |x - \bar{x}|^2 \right\}.
\]  

(12)

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

\[
\rho^+ = \arg \min_{\rho \geq 0} \left\{ a \, f (\rho) + \frac{1}{2} (\rho - \tilde{\rho})^2 \right\},
\]

(13)

\[
\phi^+ = \tilde{\phi},
\]

(14)

where \( \bar{x} = \tilde{\rho} \exp(\tilde{\phi}) \).

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 \( \bar{x} \) and its projection on the circle \( |x|^2 = d \), as illustrated in Figure 1. The position on this line varies monotonically with \( a \), so that \( \text{prox}_{a \, f} (\bar{x}) = \bar{x} \) for \( a = 0 \) and gets closer to \( \sqrt{d} \tilde{\rho} \) as \( a \) increases. The classical operator defined in (6) can thus be seen as \( \lim_{a \to \infty} \text{prox}_{a \, f} (\bar{x}) = p(x_k | d_k) \).

From this solution, we can identify three subdomains where \( \text{prox}_{a \, f} (\bar{x}) \) has different properties:

- When \( \bar{x} \in \{ x \in \mathbb{C} : |x|^2 \geq d \} \), \( \text{prox}_{a \, f} (\bar{x}) \) is single valued and thus \( f \) is \textit{prox-regular}. Furthermore, the proximity operator of \( f \) is non-expansive on this sub-domain.

- When \( \bar{x} \in \{ x \in \mathbb{C} : 0 < |x|^2 < d \} \), \( f \) is still \textit{prox-regular} but \( \text{prox}_{a \, f} \) is no longer non-expansive. Indeed, \( \| \text{prox}_{a \, f} (\bar{x}) - \text{prox}_{a \, f} (\bar{y}) \|_2^2 \geq \| \bar{x} - \bar{y} \|_2^2 \) as illustrated in Figure 2.

- When \( \bar{x} = 0 \) and \( d > 0 \), \( \text{prox}_{a \, f} \) is multivalued in 0 as all the points on the circle of radius \( \rho^+ \) are solution of (12). As a consequence, \( f \) is not \textit{prox-regular} at \( \{0\} \) and its proximity operator is not defined for this point.

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

\[
\text{prox}_{a \, f} (\bar{x}) = \begin{cases} \rho^+ & \text{if } \bar{x} = 0, \\ \rho^+ \exp(\rho^+ \phi) & \text{otherwise,} \end{cases}
\]

(15)

Let us notice that the modified projection \( P^\rho (\bar{x}, d) \) defined by Equation 8 lies also on the line between \( \bar{x} \) and its projection on the circle \( |x|^2 = d \). Its position on this line depends on the value of the relaxation parameter \( \beta \). 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 \( \sigma^2 = \text{Var} \{d\} \), the function \( f \) writes

\[
f(x) = w (|x|^2 - d)^2,
\]

(16)

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

\[
\rho^+ = \arg \min_{\rho \geq 0} \left( a \, w (\rho^2 - d)^2 + \frac{1}{2} (\rho - \tilde{\rho})^2 \right).
\]

(17)

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

\[
q_G (\rho) = \frac{d}{d \rho} \left( a \, w (\rho^2 - d)^2 + \frac{1}{2} (\rho - \tilde{\rho})^2 \right) = 4 a \, w \rho^3 + \rho (1 - 4 a \, w d) - \tilde{\rho}.
\]

(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 \( \tilde{\rho} / (4 a \, w) > 0 \). Thus, \( q_G \) has always only one positive root \( \rho^+ \). As stated in the previous section, this root must lie between \( \sqrt{d} \tilde{\rho} \) 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

\[
f(x) = |x|^2 - d \log \left( |x|^2 + b \right),
\]

(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 \( q_P (\rho) = \frac{d q_P (\rho)}{d \rho} \), with

\[
q_P (\rho) = \frac{d}{d \rho} \left( a \, f (\rho) + \frac{1}{2} (\rho - \tilde{\rho})^2 \right)
\]

\[
= (2 a + 1) \rho^3 - \tilde{\rho} \rho^2 + ((2 a + 1) b - 2 a d) \rho - b \tilde{\rho}.
\]

(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

\[
\rho^+ = \frac{\tilde{\rho} + \sqrt{8 d a (1 + 2 a) + \tilde{\rho}^2}}{2 + 4 a}.
\]

(21)
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

$$d_k = \|y_k\|^2_n + n_k,$$  \hspace{1cm} (22)

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

Then (12) becomes

$$\text{prox}_{\alpha f}(\tilde{y}) = \arg \min_{\eta \geq 0, \|u\|_2=1} \left( \alpha f(\eta) + \frac{1}{2} \|\eta u - \tilde{y}\|^2_2 \right).$$  \hspace{1cm} (23)

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

$$u^+(\eta) = \arg \min_{u, \|u\|=1} \|\eta u - \tilde{y}\|^2_2 = \frac{\tilde{y}}{\|\tilde{y}\|_2}. $$ \hspace{1cm} (24)

Thus the solution is

$$y^+ = \eta^+ \frac{\tilde{y}}{\|\tilde{y}\|_2},$$ \hspace{1cm} (25)

where $\eta^+$ is given by

$$\eta^+ = \arg \min_{\eta > 0} \left( \min_{u, \|u\|=1} \left( \alpha f(\eta) + \frac{1}{2} \|\eta u - \tilde{y}\|^2_2 \right) \right).$$ \hspace{1cm} (26)

$$= \arg \min_{\eta > 0} \left( \alpha f(\eta) + \frac{1}{2} (\eta - \|\tilde{y}\|_2)^2 \right),$$ \hspace{1cm} (27)
Fig. 9. Comparison of DR and GS performance in noisy conditions ($\sigma = 1$) using the classical projection or the proposed operator.

since

$$\min_{u, |u| = 1} \| \eta u - \hat{y} \|_2^2 = (|\eta| - \| \hat{y} \|_2)^2.$$  

(28)

Solving (27) is equivalent to solving (13) with $\bar{\rho} = \|\hat{y}\|_2$. In the case where $\|\hat{y}\|_2 = 0$ 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 $\text{prox}_f(\hat{y}) = \eta^+$ when $\|\hat{y}\|_2 = 0$. To sum up, the proximity operator for undersampled measurements is:

$$\text{prox}_f(\hat{y}) = \begin{cases} 
\eta^+, & \text{if } \|\hat{y}\|_2 = 0 \\
\eta^+ \frac{\hat{y}}{\|\hat{y}\|_2}, & \text{otherwise}.
\end{cases}$$  

(29)

This proximity operator for undersampled intensity measurements can be computed for any function $f$ that has a proximity operator in closed form such as 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 $r$ that will be estimated throughout the experiments. We computed the noisy intensities $d_A = |r| + n_A$ and $d_B = |Hr| + n_B$ at depth $z_A$ and $z_B$, where $H$ is the propagation operator from $z_B$ to $z_A$ and $n_A$ and $n_B$ are noise vectors with identical statistics given by the experimental conditions. The setup parameters are: $\lambda = 633 \text{ nm}$, pixel size: $= 5.3 \mu m$, $z_A = 1 \text{ cm}$, and $z_B = 2 \text{ cm}$.

For each experiment, we built the functions $f_{A,k}(x) = f_k(|x|^2; d_{A,k})$ and $f_{B,k}(x) = f_k(|x|^2; d_{B,k})$ according to the considered noise model. We then compared the performance of the proposed proximity operator $\text{prox}_{f_A}$ to that of the classical projection defined by (6) by estimating the complex amplitude of the wave $x^+$ 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 $x$ in plane $z_A$ is assessed by the mean of the reconstruction signal to noise ratio:

$$\text{SNR}(x) = 10 \log_{10} \frac{\|r\|_2^2}{\|r - x\|_2^2}.$$  

(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 $x^{(0)} = \sqrt{d_A}$ 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

1: procedure DR($f_A, f_B$)
2: $y^{(0)} = \sqrt{d_A}$ and $\lambda \in [0, 2]$ \quad init. ($\lambda = 1$ for all results)
3: for $n = 1, \ldots, \text{maxiter}$ do
4: \quad $x^{(n)} = \text{prox}_{\lambda f_A}(y^{(n-1)})$
5: \quad $r^{(n)} = 2x^{(n)} - y^{(n-1)}$
6: \quad $y^{(n)} = y^{(n-1)} + \lambda \left( H^T \text{prox}_{\lambda f_B}(Hr^{(n)}) - x^{(n)} \right)$
7: return $x^{(\text{maxiter})}$ \quad \triangleright 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

$$x^+ = \arg \min_{x \in \mathbb{C}^K} \left( \sum_{k=1}^{N} f_{A,k}(x_k) + \inf_{y \in \mathbb{C}^K} \left( \sum_{k'=1}^{K} f_{B,k'}(y_{k'}) + \frac{1}{2} \| Hx - y \|_2^2 \right) \right),$$  

(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 $g(x) = f(H \cdot x)$ [26]:

$$H^T H = I \quad \implies \quad \text{prox}_{xg}(x) = H^T \cdot \text{prox}_{xf}(H \cdot x),$$  

(32)

where $I$ is the identity matrix.
For the Gaussian likelihood as for the Poisson likelihood, \( f \) is not convex. The convergence of both algorithms cannot be proven 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 \( x_0 = \sqrt{d_A} \).

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 \( \alpha \). 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, \( \text{SNR}(x) \) began to worsen after some iteration, while the cost was still decreasing, as can be seen in 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:

\[
\chi^2 = \frac{1}{2K} \left( \sum_{k=1}^{K} f_A(x_y) + \sum_{k'=1}^{K} f_{b'k'}(|Hx|_{k'}) \right) < 1. \tag{33}
\]

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 \( \alpha \) has a strong effect on the speed of convergence but has little influence on the quality. However, if \( \alpha \) is too large (e.g., \( \alpha = 1 \) in Figure 11), the steps are too large and the criterion \( \chi^2 \) is well below 1 even after the first iteration. As consequence, \( \alpha \) 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 \( \alpha \) that maximize \( \text{SNR}(x^+) \).

### 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 \( \text{SNR}(x^+) \) 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 \( \text{SNR}(x^+) \) 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 (\( b_0 = 0 \)), 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 \( \text{SNR}(x^+) \), 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 \( P \), so that

\[
d_k = P \left( |x_k|^2 + b \right), \tag{34}
\]

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...
As in the previous experiments, we estimated the complex amplitude \( \hat{x} \) of the object and estimated it compared to the number of unknowns \( M \). 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 \( g_p \in \mathbb{C}^{K_1 \times K_2} \) in the detector plane \( z_0 \), we simulated \((2 \times 2)\) subsampled intensity measurements \( d_p \in \mathbb{R}^{M_1 \times M_2} \) with \( K = 2M \) using the direct model

\[
\begin{align*}
  g_p &= H_p \cdot r, \\
  d_{p,m_1,m_2} &= |s_{p,2m_1+1,2m_2}|^2 + |s_{p,2m_1+2,2m_2}|^2 + |s_{p,2m_1,2m_2+1}|^2 + |s_{p,2m_1+1,2m_2+1}|^2 + n_p
\end{align*}
\]

where \( H_p \) is the propagation operator from the plane \( z_1 \) to \( z_p \). As in the previous experiments, we estimated the complex amplitude \( x^+ \) 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 \( x^+ \) in the plane \( z_1 \) from \( P = 8 \) measurements.

The maximum-likelihood solution in this case is given by

\[
  x^+ \in \arg \min_{x \in \mathcal{C}^K} \sum_{p=1}^{P} \sum_{k=1}^{K} f_{p,k} (|H_p \cdot x|_k).
\]

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 \( z_1 = 1 \text{ cm}, z_2 = 2 \text{ cm}, z_3 = 3 \text{ cm}, z_4 = 4 \text{ cm}, z_5 = 5 \text{ cm}, z_6 = 6 \text{ cm}, z_7 = 7 \text{ cm}, \) and \( z_8 = 8 \text{ cm} \). These measurements were corrupted with additive Gaussian noise of variance \( \sigma = 0.5 \) (corresponding to SNR(\( \sigma = -2.1 \text{ dB} \)).

We have estimated the 1024 × 968 pixels complex amplitude \( x_1^+ \) 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 algo-