

# Generalized Proximal Smoothing for Phase Retrieval

MINH PHAM<sup>1</sup>, PENGHANG YIN<sup>1</sup>, ARJUN RANA<sup>2</sup>, STANLEY OSHER<sup>1</sup>,  
AND JIANWEI MIAO<sup>2</sup>

<sup>1</sup>*Department of Mathematics, University of California, Los Angeles, CA 90095, USA*

<sup>2</sup>*Department of Physics & Astronomy and California NanoSystems Institute, University of California, Los Angeles, CA 90095, USA*

<sup>1</sup>*(minhrose,yph,sjo)@math.ucla.edu*

<sup>2</sup>*(arana,miao)@physics.ucla.edu*

**Abstract:** In this paper, we report the development of the generalized proximal smoothing (GPS) algorithm for phase retrieval of noisy data. GPS is a optimization-based algorithm, in which we relax both the Fourier magnitudes and object constraints. We relax the object constraint by incorporating the Moreau-Yosida regularization and heat kernel smoothing, and derive the associated proximal mapping. We also relax the magnitude constraint into a least squares fidelity term, whose proximal mapping is available as well. GPS alternatively iterates between the two proximal mappings in primal and dual spaces, respectively. Using both numerical simulation and experimental data, we show that GPS algorithm consistently outperforms the classical phase retrieval algorithms such as hybrid input-output (HIO) and oversampling smoothness (OSS), in terms of the convergence speed, consistency of the phase retrieval, and robustness to noise.

*Index Terms*—phase retrieval, oversampling, coherent diffractive imaging, Moreau-Yosida regularization, heat kernel smoothing, primal-dual algorithm

© 2018 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

## 1. Introduction

Phase retrieval has been fundamental to several disciplines, ranging from imaging, microscopy, crystallography and optics to astronomy [1–4]. It aims to recover an object only from its Fourier magnitudes. Without the Fourier phases, the recovery can be achieved via iterative algorithms when the Fourier magnitudes are sampled at a frequency sufficiently finer than the Nyquist interval [5]. In 1972, Gerchberg and Saxton developed an iterative algorithm for phase retrieval, utilizing the magnitude of an image and the Fourier magnitudes as constraints [6]. In 1982, Fienup generalized the Gerchberg-Saxton algorithm by developing two iterative algorithms: error reduction (ER) and hybrid input-output (HIO), which use a support and positivity as constraints in real space and the Fourier magnitudes in reciprocal space [7]. In 1998, Miao, Sayre and Chapman proposed, when the number of independently measured Fourier magnitudes is larger than the number of unknown variables associated with a sample, the phases are in principle encoded in the Fourier magnitudes and can be retrieved by iterative algorithms [5]. These developments finally led to the first experimental demonstration of coherent diffractive imaging (CDI) by Miao and collaborators in 1999 [8], which has stimulated wide spread research activities in phase retrieval, CDI, and their applications in the physical and biological sciences ever since [2, 9, 10].

For a finite object, when its Fourier transform is sampled at a frequency finer than the Nyquist interval (i.e. oversampled), mathematically it is equivalent to padding zeros to the object in real space. In another words, when the magnitude of the Fourier transform is oversampled, the correct phases correspond to the zero-density region surrounding the object, which is known as the oversampling theorem [5, 11]. The phase retrieval algorithms iterate between real and reciprocal space using zero-density region and the Fourier magnitudes as dual-space constraints. A support is typically defined to separate the zero-density region from the object. The positivity constraint

is applied to the density inside the support. In the ER algorithm, the no-density region outside the support and the negative density inside the support are set to zero in each iteration [7]. The HIO algorithm relaxes the ER in the sense that it gradually reduces the densities that violate the object constraint instead of directly forcing them to zero [7]. This relaxation often leads to good reconstructions from noise-free patterns. However, in real experiments, the diffraction intensities, which are proportional to the square of Fourier magnitudes, are corrupted by a combination of Gaussian and Poisson noise and missing data. In the presence of experimental noise and missing data, phase retrieval becomes much more challenging, and the ER and HIO algorithms may only converge to sub-optimal solutions. Simply combining ER and HIO still suffers from stagnation and the iterations can get trapped at local minima [12]. To alleviate these problems, more advanced phase retrieval algorithms have been developed such as the shrink-wrap algorithm and guided HIO (gHIO) [4, 13]. In 2010, a smoothness constraint in real space was first introduced to improve the phase retrieval of noisy data [14]. Later, a noise robust framework was implemented for enhancing the performance of existing algorithms [15]. Recently, Rodriguez et al. proposed to impose the smoothness constraint on the no-density region outside the support by applying Gaussian filters [16]. The resulting oversampling smoothness (OSS) algorithm successfully reduces oscillations in the reconstructed image, and is more robust to noisy data than the existing algorithms.

Since phase retrieval can be cast as a non-convex minimization problem, many efforts have been made to study phase retrieval algorithms from the viewpoint of optimization. For example, Bauschke et al. [17] related HIO to a particular relaxation of the Douglas-Rachford algorithm [18] and introduced the hybrid projection reflection algorithm [17, 19]. Using similar ideas, researchers further proposed several projection algorithms such as iterated difference map [20] and relaxed averaged alternation reflection [21]. In [22], Chen and Fannjiang analyzed a Fourier-domain Douglas-Rachford algorithm for phase retrieval. By taking noise into account, the Wirtinger Flow [23] relaxes the magnitude constraint into a fidelity term that measures the misfit of Fourier data, to which Wirtinger gradient descent is applied. Other methods in this line include alternating direction methods [24–26] that have been widely used in image processing, as well as lifting approaches [27] such as PhaseLift [28, 29] by Candès et al. and its variants [30, 31].

In this paper, we propose an optimization-based phase retrieval method, termed generalized proximal smoothing (GPS), which effectively addresses the noise in both real and Fourier spaces. Motivated by the success of OSS [16], GPS incorporates the idea of Moreau-Yosida [32, 33] regularization with heat kernel smoothing, to relax the object constraint into a continuous regularizer. We further relax the magnitude constraint into a least squares fidelity term, for de-noising in Fourier space. To minimize the resulting primal-dual formulation, GPS iterates back and forth between efficient proximal mappings of the two relaxed functions, respectively. Our experimental results using noisy experimental data of biological and inorganic specimens demonstrate that GPS consistently outperforms the state-of-the-art algorithms HIO and OSS in terms of both speed and robustness. We also refer readers to the recent paper [34] about training quantized neural networks, which shows another success of using Moreau-Yosida regularization to relax the hard constraint.

**Notations.** Let us fix some notations. For any complex-valued vectors  $u, v \in \mathbb{C}^n$ ,  $\bar{u}$  is the complex conjugate of  $u$ , whereas  $u^* := \bar{u}^\top$  is the Hermitian transpose.  $\text{Re}(u)$  and  $\text{Im}(u)$  are the real and imaginary parts of  $u$ , respectively.

$$\langle u, v \rangle := u^* v = \sum_{i=1}^n \bar{u}_i v_i \text{ is the Hermitian inner product of } u \text{ and } v. \quad u \odot v \text{ is the element-wise}$$

(Hadamard) product of  $u$  and  $v$  given by  $(u \odot v)_i = u_i v_i$ .  $\|u\| := \sqrt{\langle u, u \rangle}$  denotes the  $\ell_2$  norm of  $u$ . Given any Hermitian positive semi-definite matrix  $K \in \mathbb{C}^{n \times n}$ , we define  $\|u\|_K := \sqrt{\langle u, K u \rangle}$ .

$\arg(u) := \tan^{-1} \left( \frac{\text{Im}(u)}{\text{Re}(u)} \right)$  denotes the argument (or phase) of a non complex-valued vector  $u = \text{Re}(u) + \mathbf{i} \cdot \text{Im}(u)$ . For convenience we set  $\arg(0) = 1$ .

$\mathcal{I}_X$  is the characteristic function of a closed set  $X \subset \mathbb{C}^n$  given by  $\mathcal{I}_X(x) = 0$  if  $x \in X$  and  $\mathcal{I}_X(x) = \infty$  otherwise

$\text{proj}_X(u) := \arg \min_{v \in X} \|v - u\|$  is the projection of  $u$  onto  $X$ ,

$\text{prox}_f(u) := \arg \min_v \left\{ f(v) + \frac{1}{2} \|v - u\|^2 \right\}$  is the proximal mapping of the function  $f$

$h^*(y) := \sup_{u \in \mathbb{C}} \{ \text{Re}\langle y, u \rangle - h(u) \}$  is the extended Legendre transformation to the complex plane of a real-valued function  $h$

## 2. The Proposed Model

First, we recall the phase retrieval problem setting. Consider the reconstruction of a 2D image  $u$  defined on a discrete lattice  $\Omega := \{(i, j) : 1 \leq i \leq n_1, 1 \leq j \leq n_2\}$ .

For simplicity, we represent  $u$  in terms of a vector in  $\mathbb{R}^n$  by the lexicographical order with  $n = n_1 \times n_2$ . Then  $u_i$  represents the density of image at the  $i$ -th pixel. Due to oversampling, the object densities reside in a sub-domain  $S \subset \Omega$  known as the support, and  $u$  is supposed to be zero outside  $S$ . Throughout the paper, we assume that the support  $S$  is centered around the domain  $\Omega$ . The object constraint is  $\mathcal{S} := \{u \in \mathbb{R}^n : u_i \geq 0 \text{ if } i \in S, u_i = 0 \text{ otherwise}\}$ .

The Fourier magnitude data is obtained as  $b = |\mathcal{F}u|$ , where  $\mathcal{F} : \mathbb{R}^{n_1 \times n_2} \rightarrow \mathbb{C}^{n_1 \times n_2}$  is the discrete Fourier transform (DFT). We denote the magnitude constraint by  $\mathcal{T} := \{u \in \mathbb{R}^n : |\mathcal{F}u| = b\}$ .

In the absence of noise, phase retrieval (PR) problem is simply to

$$\text{find } u \in \mathbb{R}^n, \quad \text{such that } u \in \mathcal{S} \cap \mathcal{T}. \quad (1)$$

This amounts to the following composite minimization problem

$$\min_{u \in \mathbb{R}^n} f(u) + g(\mathcal{F}u), \quad (2)$$

where  $f(u) := \mathcal{I}_S(u)$  and  $g(z) := \mathcal{I}_{|z|=b}(z)$  are two characteristic functions that enforce the object and Fourier magnitudes constraints. Note that  $f$  is a closed and convex function while  $g$  is closed but non-convex, which give rise to the non-convex optimization problem of (2).

### 2.1. A new noise-removal model

In real experiments, the Fourier data are contaminated by experimental noise. Moreover, the densities outside the support are not exactly equal to zero either. In the noisy case, the image to be reconstructed no longer fulfills either the Fourier magnitudes or the object constraint. The ER algorithm, which alternatively projects between these two constraints, apparently does not take care of the noise. The HIO “relaxes” the object constraint on densities wherever it is violated. This relaxation only helps in the noiseless case. In the presence of noise, the feasible set  $\mathcal{S} \cap \mathcal{T}$  can be empty, and alternating projection methods like ER and HIO may fail to converge and keep oscillating. The OSS [16] improves the HIO by applying extra Gaussian filters to smooth the densities outside the support at different stages of the iterative process. None of them, however, seems to properly address the corruption of the Fourier magnitudes.

Introducing the splitting variable  $z = \mathcal{F}u \in \mathbb{C}^n$ , we reformulate (2) as

$$\min_{u, z \in \mathbb{C}^n} f(u) + g(z) \quad \text{subject to} \quad z = \mathcal{F}u. \quad (3)$$

In the presence of noise, we seek to relax the characteristic functions  $f$  and  $g$  that enforce hard constraint into soft constraints. We first relax  $g$  into a least square fidelity

$$g_\sigma(z) = \frac{1}{2\sigma} \| |z| - b \|^2. \quad (4)$$

This fidelity term has been considered in the literature by assuming the measurements being corrupted by i.i.d. Gaussian noise; see [24] for example. In practice, we observe that it works well even with a combination of Gaussian and Poisson noise. This relaxation is related to the Moreau-Yoshida regularization [32, 33] and will be explained in detail in the appendix.

Following this line, we further relax  $f = \mathcal{I}_S$  into

$$f_G(u) := \inf_v \left\{ f(v) + \frac{1}{2} \|v - u\|_{G^{-1}}^2 \right\} = \inf_{v \in S} \frac{1}{2} \|v - u\|_{G^{-1}}^2 \quad (5)$$

for some Hermitian positive definite matrix  $G$ . The choice of  $G$  here is tricky, and will be discussed later in section 3 and appendix. The relaxation of both constraints thus leads to the proposed noise-removal model

$$\min_{u, z \in \mathbb{C}^n} f_G(u) + g_\sigma(z) \quad \text{subject to} \quad z = \mathcal{F}u. \quad (6)$$

This constraint optimization can be solved by a splitting method, such as alternating direction method of multipliers (ADMM) [35–37] or primal dual hybrid gradient (PDHG) algorithm [38–41].

## 2.2. A primal-dual formulation

We solve the constraint optimization problem (6) by introducing the Lagrangian

$$\mathcal{L}(u, z; y) = f_G(u) + g_\sigma(z) + \text{Re}\langle y, \mathcal{F}^* z - u \rangle, \quad (7)$$

where  $\mathcal{F}^* = \mathcal{F}^{-1}$  is the adjoint of  $\mathcal{F}$  or the inverse DFT. The corresponding Karush-Kuhn-Tucker (KKT) condition is

$$y \in \partial f_G(u), \quad -\mathcal{F}y \in \partial g_\sigma(z) \iff \mathcal{F}^* z = u \in \partial f_G^*(y), \quad -\mathcal{F}y \in \partial g_\sigma(z), \quad (8)$$

where the right hand side equivalence is obtained by the convex conjugate property and is exactly the KKT condition of the following min-max saddle point problem

$$\min_z \max_y g_\sigma(z) - f_G^*(y) + \text{Re}\langle z, \mathcal{F}y \rangle, \quad (9)$$

with  $f_G^*(y)$  is the generalized Legendre-Fenchel transformation of  $f_G^*$  and has an explicit form thanks to Moreau-Yoshida decomposition

$$f_G^*(y) = f^*(y) + \frac{1}{2} \|y\|_G^2, \quad (10)$$

where  $f^*$  is the Legendre-Fenchel transformation of  $f$  defined as

$$f^*(y) = \sup_{v \in S} \text{Re}\langle y, v \rangle = \begin{cases} 0 & \text{if } \text{Re}(y) \leq 0 \text{ on } S, \\ \infty & \text{otherwise.} \end{cases} \quad (11)$$

For notational convenience, we denote  $\mathcal{S}^* := \{y \in \mathcal{S}^n : \text{Re}(y) \leq 0 \text{ on } S\}$  the dual constraint set of  $\mathcal{S}$ . Our finalized min-max problem is thus given by

$$\min_z \max_{y \in \mathcal{S}^*} \frac{1}{2\sigma} \| |z| - b \|^2 + \frac{1}{2} \|y\|_G^2 + \text{Re}\langle z, \mathcal{F}y \rangle. \quad (12)$$

### 3. Generalized Proximal Smoothing algorithm

We carry out the minimization of the saddle point problem (9) or (12) by a generalized primal dual hybrid gradient (PDHG) algorithm [38–41], which iterates

$$\begin{cases} z^{k+1} = \text{prox}_{tg_\sigma}(z^k - t\mathcal{F}y^k) \\ y^{k+1} = \text{prox}_{sf_G^*}(y^k + s\mathcal{F}^{-1}(2z^{k+1} - z^k)) \end{cases} \quad (13)$$

for some step sizes  $s, t > 0$ . The update of  $z^{k+1}$  calls for computing the proximal mapping of  $tg_\sigma$  [24], whose analytic solution is given by

$$\text{prox}_{tg_\sigma}(z) = \arg \min_{v \in \mathbb{C}^n} \frac{1}{2\sigma} \| |v| - b \|^2 + \frac{1}{2t} \| v - z \|^2 = \frac{b \odot \exp(\mathbf{i} \cdot \arg(z)) + (\sigma/t)z}{1 + \sigma/t}, \quad (14)$$

which is essentially a weighted average between  $z$  and its projection onto the magnitude constraint  $\{z \in \mathbb{C}^n : |z| = b\}$ .

Moreover, we need to find the proximal mapping of  $sf_G^*$  for updating  $y^{k+1}$ , which reduces to

$$\begin{aligned} \text{prox}_{sf_G^*}(y) &= \arg \min_{v \in \mathbb{C}^n} f_G^*(v) + \frac{1}{2s} \| v - y \|^2 = \arg \min_{v \in \mathbb{C}^n} f^*(v) + \frac{1}{2} \| v \|_G^2 + \frac{1}{2s} \| v - y \|^2 \\ &= \arg \min_{v \in \mathcal{S}^*} \frac{1}{2} \| v \|_G^2 + \frac{1}{2s} \| v - y \|^2. \end{aligned} \quad (15)$$

Eq. (15) has closed-form solution only when  $G$  is a diagonal matrix. In the other case, we provide an approximation.

We devise two versions of GPS algorithm based on different choices of  $G$ .

#### 3.1. Real space smoothing

One choice of  $G$  is  $G = \gamma D^\top D$ . Here  $D$  is the discrete gradient operator, and then  $D^\top D$  is the negative of discrete Laplacian. In this case,

$$f_G^*(y) = f^*(y) + \frac{\gamma}{2} \| Dy \|^2, \quad (16)$$

which we shall refer as the real space smoothing. Since  $G$  is not diagonal, the closed-form solution to Eq. (15) is not available. For small  $\gamma$ , we approximate the solution by 2 steps: projection and smoothing,

$$\text{prox}_{sf_G^*}(y) \approx (I + s\gamma D^\top D)^{-1} \text{proj}_{\mathcal{S}^*}(y). \quad (17)$$

The linear inversion is in fact the Laplacian smoothing that follows the projection to ensure the smoothness of the reconstructed image after each iteration. The real space smoothing is related to the diffusion process and can be approximated by a low-pass filter. Recall  $\mathcal{G}_t(x) := \frac{1}{(4\pi t)^{n/2}} \exp(-\frac{\|x\|^2}{4t})$  is a (heat) Gaussian kernel and its Fourier transform  $W_t(\xi) = \exp(-4\pi^2 t |\xi|^2)$  is a normalized Gaussian function. Then replacing the linear inversion by the Gaussian kernel and convolution give a fast approximated implementation of Eq. (17) for small  $\gamma$

$$y^{k+1} = \mathcal{G}_{s\gamma} * \text{proj}_{\mathcal{S}^*}(y^k + s\mathcal{F}^{-1}(2z^{k+1} - z^k)), \quad (18)$$

The projection on  $\mathcal{S}^*$  can be computed as follows

$$\text{proj}_{\mathcal{S}^*}(y)_i = \begin{cases} \text{Re}(y_i)^- + \mathbf{i} \cdot \text{Im}(y_i) & \text{if } i \in \mathcal{S}, \\ y_i & \text{otherwise.} \end{cases} \quad (19)$$

Here  $x^- := \min(0, x)$  for  $x \in \mathbb{R}$ . The remaining convolution can be done via the efficient DFT via the computation

$$\mathcal{G}_\gamma * u = \text{Re} \left\{ \mathcal{F}^{-1} (W_\gamma \odot \mathcal{F}u) \right\}. \quad (20)$$

Using Eq. (14), Eq. (19), and Eq. (20), we arrive at our first algorithm

Algorithm 1: GPS-R features low-pass filters for smoothing. Here we abuse notation  $\gamma$  to imply the filter. Inspired by OSS [16], we choose an increasing sequence of spatial frequency filters  $\{\gamma_l\}$  (a sequence of finer filters). In our experiments, we do 1000 iterations with totally 10 stages. Each stage contains 100 iterations, in which we stick with the same filter frequency. We monitor the R-factor (relative error) during the iterative process, which is defined as

$$R_F(u^k) = \frac{\sum_i |\mathcal{F}u^k|_i - b_i}{\sum_i b_i}. \quad (21)$$

The reconstruction with minimal  $R_F$  at each state is fed into the next stage. By applying the smoothing on the entire domain, GPS-R can remove noise in real space and obtain the spatial resolution with fine features.

---

**Algorithm 1** GPS-R: GPS with smoothing in real space.

---

**Input:** measurements  $b$ , regularization parameters  $\{\gamma_l\}_{l=1}^{10}$ , step sizes  $s, t > 0$

**Initialize:**  $z^0, y^0$ .

$R_F^{best} = 1, z_{best} = z^0$

**for**  $l = 1, \dots, 10$  **do**

$y^0 = y_{best}, z^0 = z_{best}$

**for**  $k = 1, \dots, 100$  **do**

$z^{k+1} = \text{prox}_{t\mathcal{G}_\sigma}(z^k - t\mathcal{F}y^k)$

$y^{k+1/2} = \text{proj}_{\mathcal{S}^*}(y^k + s\mathcal{F}^{-1}(2z^{k+1} - z^k))$

$y^{k+1} = \mathcal{G}_{\gamma_l} * y^{k+1/2}$

**if**  $R_F^k < R_F^{best}$ , **then**  $R_F^{best} = R_F^k, z_{best} = z^k$ , **end if**

**end for**

**end for**

**Output:**  $\mathcal{F}^{-1} z_{best}$

---

### 3.2. Fourier space smoothing

Another simple choice of  $G$  is the diagonal matrix  $G = \gamma \text{Diag}(r \odot r)$  where  $r \in \mathbb{R}^n$  and  $r_i$  is the distance in the original 2D lattice between the  $i$ -th pixel and the center of image. Note that  $G$  is not invertible since  $r_i = 0$  for the pixel at the center. By Eq. (41),

$$f_G(u) = \begin{cases} \sum_i \frac{|u - \text{proj}_{\mathcal{S}}(u)|_i^2}{2\gamma r_i^2} & \text{if } u_i \geq 0 \text{ at the center,} \\ \infty & \text{otherwise.} \end{cases} \quad (22)$$

So  $f_G(u)$  is a weighted sum of squares penalty on  $u$ . The weight is inversely proportional to the squared radius, which is infinity for density in the center. The further the density off the center, the smaller the penalty for the object constraint being violated.

By the Parseval's identity, for square-integrable function  $u$ , we have

$$\int \left| \frac{d}{d\xi} \hat{u}(\xi) \right|^2 d\xi = \int |x u(x)|^2 dx, \quad (23)$$

where  $\hat{u}$  is the Fourier transform of  $u$ . In the discrete setting, this amounts to

$$\|D\mathcal{F}u\|^2 = \|r \odot u\|^2, \quad (24)$$

Therefore, by Eq. (10),

$$f_G^*(y) = f^*(y) + \frac{\gamma}{2} \|r \odot y\|^2 = f^*(y) + \frac{\gamma}{2} \|D\mathcal{F}y\|^2. \quad (25)$$

This means that we are smoothing  $f^*(y)$  by regularizing with the  $\ell_2$  gradient of Fourier coefficients of  $y$ . We thus refer it as the Fourier space smoothing.

Since  $G$  is diagonal, Eq. (15) has the closed-form solution

$$\text{prox}_{sf_G^*}(y) = \frac{1}{1 + s\gamma r^2} \odot \text{proj}_{S^*}(y) \approx \exp(-s\gamma r^2) \odot \text{proj}_{S^*}(y). \quad (26)$$

Define  $W_{\frac{s\gamma}{4\pi^2}} := \exp(-s\gamma r^2)$ . The above quantity can be approximated by a direct multiplication for small  $\gamma$ . We update  $y^{k+1}$  as

$$y^{k+1} = W_{\frac{s\gamma}{4\pi^2}} \odot \text{proj}_{S^*}(y^k + s\mathcal{F}^{-1}(2z^{k+1} - z^k)). \quad (27)$$

GPS with smoothing in Fourier space (GPS-F) is summarized in Algorithm 2.

---

**Algorithm 2** GPS-F: GPS with smoothing in Fourier space.

---

**Input:** measurements  $b$ , regularization parameters  $\{\gamma_l\}_{l=1}^{10}$ , step sizes  $s, t > 0$ .

**Initialize:**  $z^0, y^0$ .

$R_F^{best} = 1, z_{best} = z^0$

**for**  $l = 1, \dots, 10$  **do**

$y^0 = y_{best}, z^0 = z_{best}$

**for**  $k = 1, \dots, 100$  **do**

$z^{k+1} = \text{prox}_{tg_\sigma}(z^k - t\mathcal{F}y^k)$

$y^{k+1/2} = \text{proj}_{S^*}(y^k + s\mathcal{F}^{-1}(2z^{k+1} - z^k))$

$y^{k+1} = W_{\frac{s\gamma_l}{4\pi^2}} \odot y^{k+1/2}$

**if**  $R_F^k < R_F^{best}$ , **then**  $R_F^{best} = R_F^k, z_{best} = z^k$ , **end if**

**end for**

**end for**

**Output:**  $\mathcal{F}^{-1}z_{best}$

---

### 3.3. Incomplete measurements

In practice, not all diffraction intensities can be experimentally measured. For example, to prevent a detector from being damaged by an intense X-ray beam, either a beamstop has to be placed in front of the detector to block the direct beam or a hole has to be made at the center of the detector, resulting in missing data at the center [42]. Furthermore, missing data may also be present in the form of gaps between detector panels. For incomplete data, the alternating projection algorithms skip the projection onto the magnitude constraint in this region. Similarly, we only apply the relaxation  $g_\sigma = \frac{1}{2\sigma} \sum_i ||z_i| - b_i|^2$  on the known data for GPS. A simple exercise shows that

$$z_i^{k+1} = \begin{cases} \left( \text{prox}_{tg_\sigma}(z^k - t\mathcal{F}y^k) \right)_i & \text{if } b_i \text{ is known,} \\ (z^k - t\mathcal{F}y^k)_i & \text{otherwise.} \end{cases} \quad (28)$$

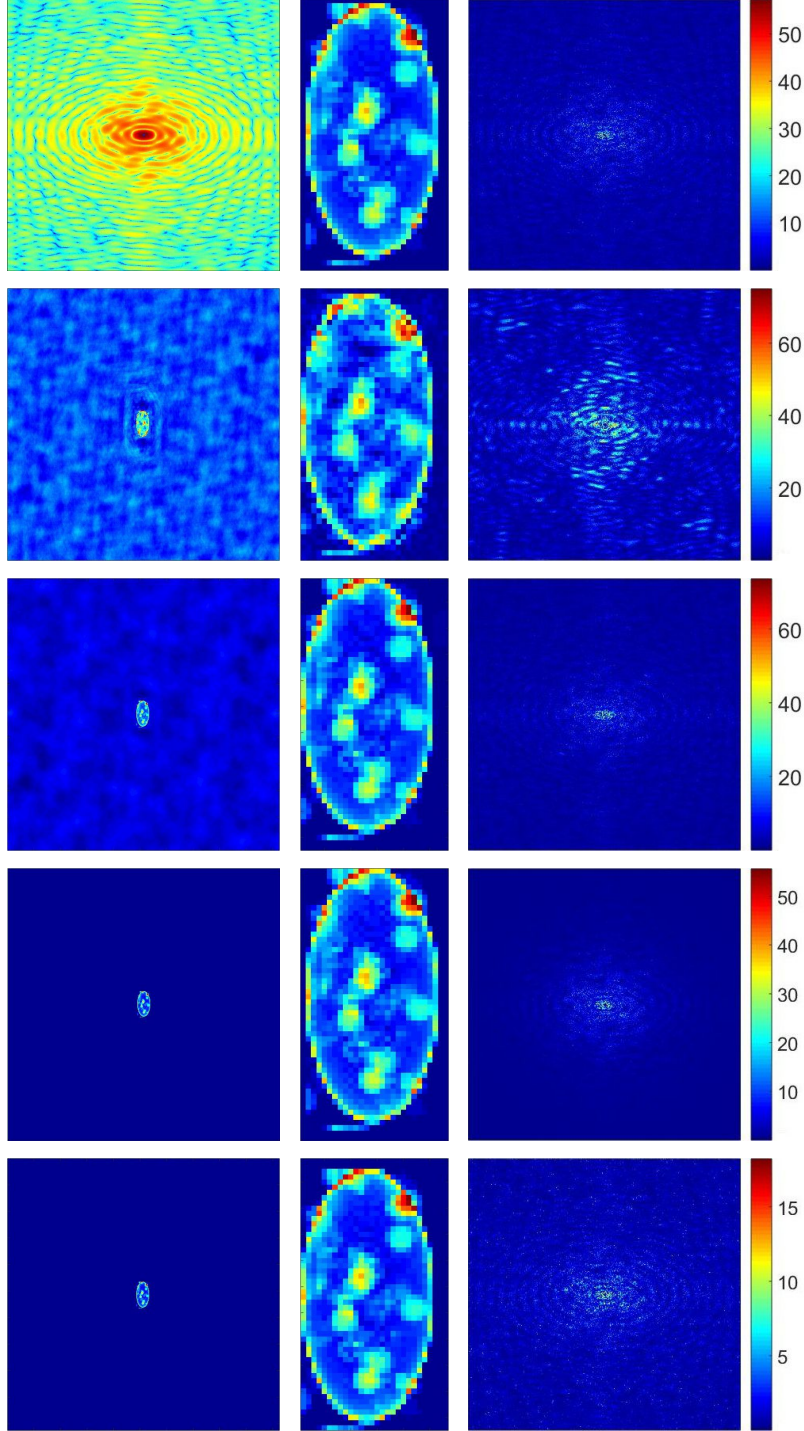


Fig. 1: First row: vesicle model with log-scaled diffraction pattern (left) , zoom-in image (center) and residual (right). Second row: HIO:  $R_F = 12.87\%$ ,  $R_{real} = 21.14\%$ . Third row: OSS:  $R_F = 6.08\%$ ,  $R_{real} = 3.59\%$ . Forth row: GPS-R  $R_F = 5.90\%$ ,  $R_{real} = 2.85\%$ . Fifth row: GPS-F  $R_F = 5.89\%$  and  $R_{real} = 0.7\%$ . Third column: residual.



## 4. Experimental Results

### 4.1. Reconstruction from simulated data

We expect GPS to be a reliable PR algorithm in the reconstruction of the images of weakly scattering objects, in particular biological specimens, which have become more popular [43]. Since OSS has been shown to consistently outperform ER, HIO, ER-HIO, NR-HIO [16], we perform both quantitative and qualitative comparisons between GPS and OSS. To simulate realistic experimental conditions, the Fourier magnitudes of a vesicle model are first corrupted with 5% Poisson noise.  $R_{noise}$  is defined to be the relative error with respect to the noise-free Fourier measurements

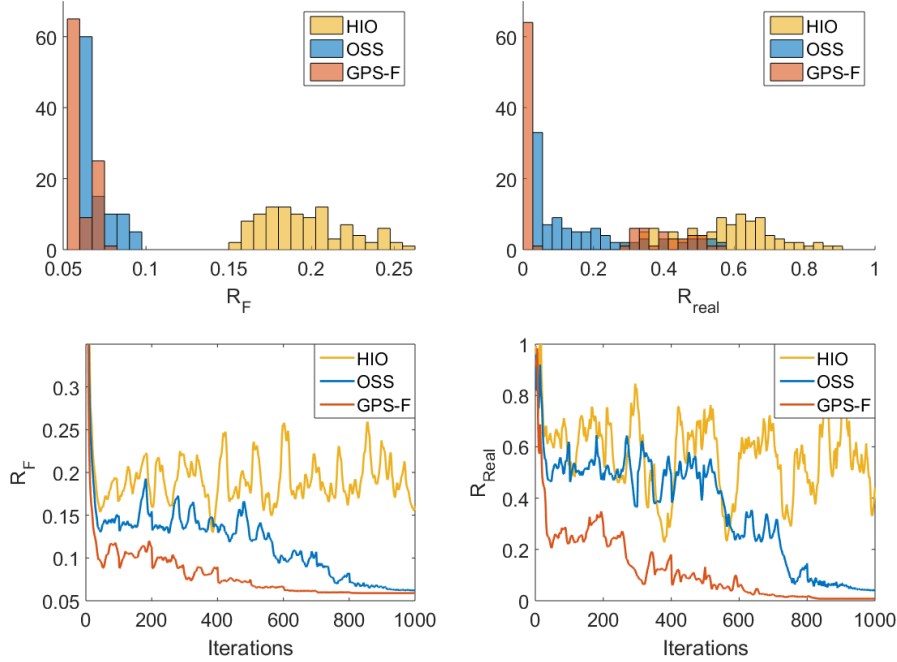


Fig. 2: Histogram (first row), and convergence (second row) of  $R_F$  and  $R_{real}$  on Vesicle data using HIO, OSS, GPS. GPS consistently produces smaller  $R_F$  and  $R_{real}$  than HIO or OSS. Moreover, GPS converges fastest with the fewest oscillations.

$$R_{noise} = \frac{\sum_i |b_i - |\mathcal{F}u^o||}{\sum_i |\mathcal{F}u^o|_i}, \quad (29)$$

where  $u^o$  is the noise-free model, and  $b$  are noisy Fourier magnitudes. Due to the discrete nature of photon counting, experimentally measured data inherently contain Poisson noise that is directly related to the incident photon flux on the object. In addition to Poisson noise, the data is also corrupted with zero-mean Gaussian noise to simulate readout from a CCD. Any resulting negative values are set to zero. Therefore, an accurate simulation of  $b_i$  can be calculated as

$$b_i = \sqrt{\text{Poisson}\left(\frac{|\mathcal{F}u^o|_i^2 \cdot \text{flux}}{\|\mathcal{F}u^o\|^2}\right) + \mathcal{N}(0, \sigma)}, \quad (30)$$

where  $\sigma$  is proportional to the readout noise. [44]

In some cases, the reconstructed image by an algorithm yields a small relative error  $R_F$  but has low quality. This is the issue of over-fitting, an example of which can be seen in certain reconstructions using ER-HIO [16]. Smoothing is a technique to avoid data over-fitting. To validate results and show that our algorithm does not develop over-fitting, we measure the difference between the reconstructed image and the model by

$$R_{real} = \frac{\sum_i |u_i^k - u_i^o|}{\sum_i u_i^o}. \quad (31)$$

In addition, we also look at the residual  $\text{Res} = \|\mathcal{F}u\| - b\|$  which measures the difference between the Fourier magnitudes of the reconstructed image and the experimental measurements. The residual can validate the noise-removal model, telling how much noise is removed. Figure 1 shows the reconstruction of vesicle model from simulated noisy data using HIO, OSS, GPS-R, and GPS-F. GPS-F and GPS-R obtain lower  $R_F$  and  $R_{real}$  than OSS. Moreover, GPS-F can get very close to the ground truth with  $R_F = 5.90\%$  and  $R_{real} = 0.7\%$ . In addition to lower  $R$  values, GPS-R and GPS-F converge to zero outside the support. They both obtain lower residuals than OSS, specifically GPS-F produces the least residual. If we choose larger parameter for the  $\ell_2$  gradient regularizer in Fourier space, we will get a smoother residual. Overall for realistic Gaussian and Poisson noise in Fourier space, GPS-F is a suitable noise-removal model.

Figure 2 shows the histogram and the convergence of  $R_F$  and  $R_{real}$  on 100 independent, randomly seeded runs using HIO, OSS and GPS on the simulated vesicle data. The histogram shows that GPS is more consistent and robust than OSS. It has a higher chance to converge to good minima with lower  $R_F$  and  $R_{real}$  than OSS. Furthermore,  $R_F$  and  $R_{real}$  of OSS scatter widely due to the initial value dependency. In contrast, GPS is more selective and less dependent on initial values.  $R_F$  and  $R_{real}$  of GPS are seen at a lower mean minimum with less variance.

Similar to HIO and ER, OSS keeps oscillating until a finer low-pass filter is applied. In contrast, GPS converges faster and is less oscillatory than OSS. In the presence of noise, alternating projection methods (ER, HIO, OSS) keep oscillating but do not converge. Applying smoothing and replacing the measurement constraint by the least squares fidelity term  $g_\sigma(z) = \frac{1}{2\sigma} \| |z| - b \|^2$  helps to reduce the oscillations; hence, the method can converge to a stationary point. Note that larger  $\sigma$  reduces more oscillations, but also decreases the chance to escape from local minima. Alternating projection methods have  $\sigma = 0$  since they impose measurement constraints. GPS obtains both smaller  $R_F$ ,  $R_{real}$ , and lower variance. Even though  $R_F$  are close to each other,  $R_{real}$  of GPS is much smaller than OSS. This means GPS recovers the vesicle cell with higher quality than OSS. This simulation shows that GPS is more reliable and consistent than OSS.

## 4.2. Reconstructions from experimental data

### 4.2.1. S. Pombe Yeast Spore

To demonstrate the applicability of GPS to biological samples, we do phase retrieval on the diffraction pattern in figure 3 taken of a *S. Pombe* yeast spore from an experiment done using beamline BL29 at SPring-8 [45]. We do 500 independent, randomly seeded reconstructions with each algorithm and record  $R_F$ , excluding the missing center. We choose default parameters for these experiments:  $t = 1$ ,  $s = 0.9$ . The sequence of low-pass filters are chosen to be the same as in OSS [16]. For the first 400 iterations,  $\sigma = 0.01$ , then is increased to  $\sigma = 0.1$  for the remaining 600 iterations. The left column of figure 3 is the mean of the best 5 reconstructions obtained by the respective algorithm. The right column shows the variance of the same 5 reconstructions. It is evident from the variance that GPS achieves more consistent reconstructions. Figure 4 shows the histogram and convergence of  $R_F$ . We can conclude that not only are GPS-R results the most consistent, but also the most faithful to the experimental data.

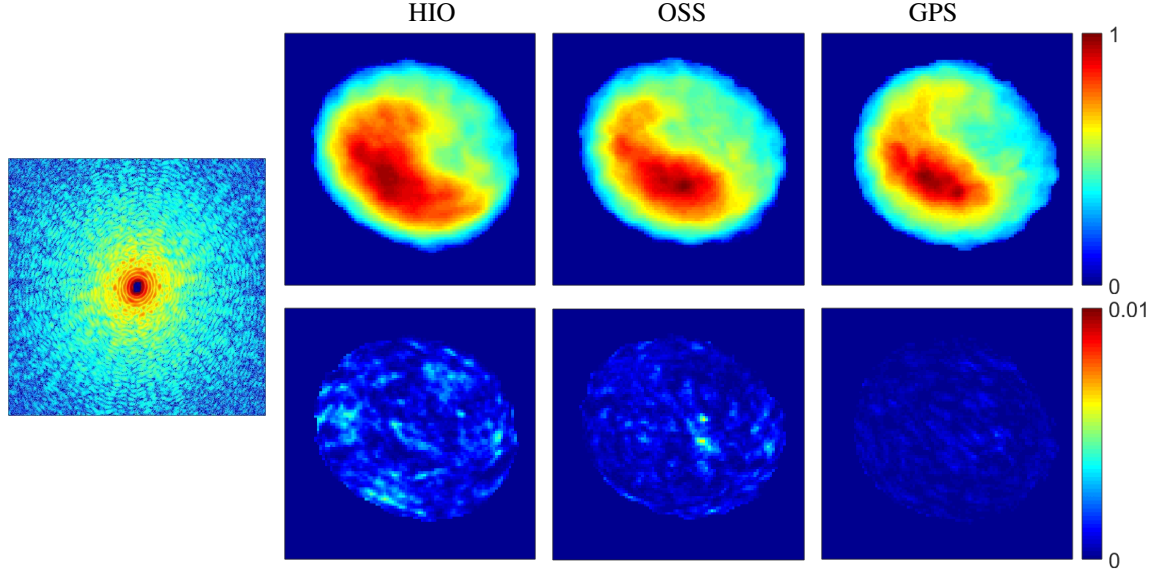


Fig. 3: *S. pombe* yeast spore log-scaled diffraction pattern, size  $500 \times 500$  (top). Means (first row) and variance (second row) are computed from the best 5 of 500 independent reconstructions. HIO:  $R_F = 15.697\% \pm 0.526\%$ , OSS:  $R_F = 9.775\% \pm 0.202\%$ , and GPS:  $R_F = 8.672\% \pm 0.025\%$ .

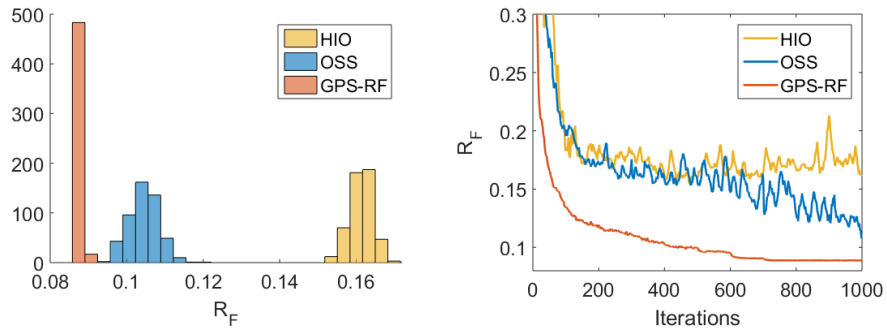


Fig. 4: Top: histogram of  $R_F$  in 500 independent runs (top). Bottom: the convergence curves of a single construction of  $R_F$  on *S. pombe* yeast spore by GPS-RF, OSS, and HIO.

#### 4.2.2. Nanorice

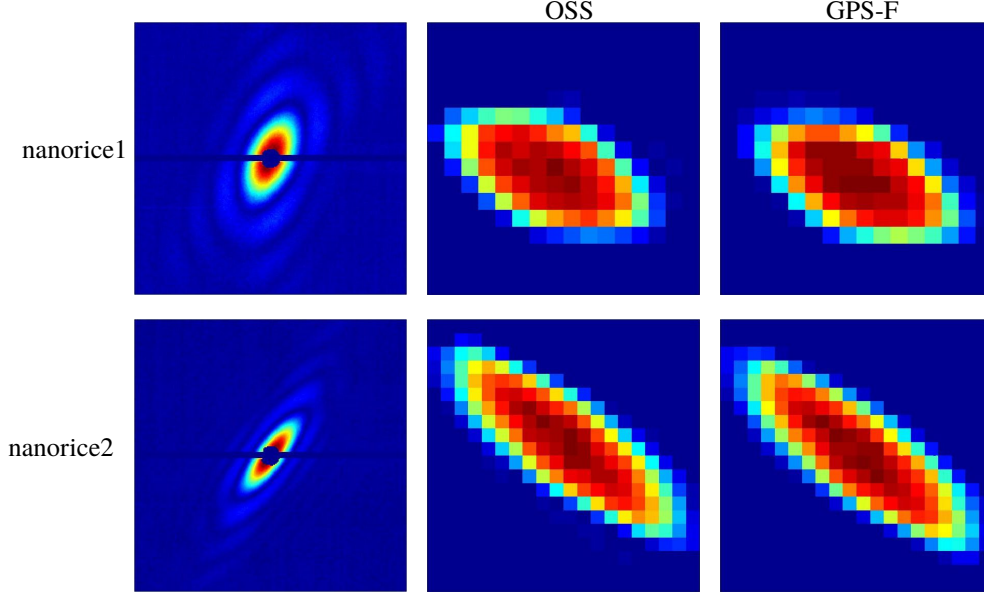


Fig. 5: Diffraction pattern of nanorice1 and nanorice2  $253 \times 253$  (first column) and reconstructions using OSS:  $R_F = 18.23\%$ ,  $16.32\%$  and GPS-F:  $R_F = 17.40\%$ ,  $15.83\%$  respectively. GPS obtains less noise on the boundary and lower relative error  $R_F$ .

To demonstrate the generality of GPS, we also do testing with experimental data of inorganic samples. The diffraction patterns shown in the top row of figure 5 from ellipsoidal iron oxide nanoparticles (referred to as ‘nanorice1’ and ‘nanorice2’) were taken at the AMO instrument at LCLS at an X-ray energy of 1.2keV [46]. This data is freely available online on the CXIDB [47]. We choose default parameters for these experiments:  $t = 1$ ,  $s = 0.9$ . The sequence of low-pass filters are chosen to be the same as in OSS [16]. The fidelity parameter  $\sigma$  is chosen small for the first 800 iterations, specifically  $\sigma \in [0, 0.01]$ , to produce oscillations which is necessary for the algorithm to skip bad local minima. Once the reconstruction arrives at a good local minimum region, we increase  $\sigma$  to reduce oscillations. This later value of  $\sigma$  depends on noise level and data. We test different values of  $\sigma$  and  $\sigma = 1$  has been found to be the optimal for both nanorice data. Figure 5 shows OSS(second row) and GPS-F(third row) reconstructions of the two nanorice particles. Figure 6 shows again that GPS obtains more consistent and faithful reconstructions as compared to those obtained by OSS. GPS-F with  $\sigma = 1$  converges to lower relative error than OSS at all times. OSS cannot get lower relative error because  $\sigma = 0$  does not work for this case. In general, alternating projection methods do not treat noise correctly. For example, in this case, HIO keeps oscillating but does not converge. Therefore, its results are omitted here. The better approach, OSS model, can reduce oscillations by smoothing but this is not enough. In contrast, the least squares  $g_\sigma(z) = \frac{1}{2\sigma} |||z| - b||^2$  of GPS works for noise removal since relaxing the constraints allows GPS to reach lower relative error. The values of  $\sigma$  depend on noise level and type. To optimize the convergence of GPS-F on nanorice2, we apply  $\sigma = 0.01$  for the first 400 iterations,  $\sigma = 0.1$  for the next 300 iterations, and  $\sigma = 1$  for the last 300 iterations. This test shows the effect of  $\sigma$  on the convergence. OSS ( $\sigma = 0$ ) oscillates the most. GPS with  $\sigma = 0.01, 0.1, 1$  oscillates less and less. As  $\sigma$  increases, GPS also gets to lower  $R_F$ . The algorithm finally reaches a stable minimum as  $\sigma$  goes up to 1. Continuing to increase  $\sigma$  does not help with  $R_F$ . Choosing large  $\sigma$  in the beginning may reduce oscillations but also limit

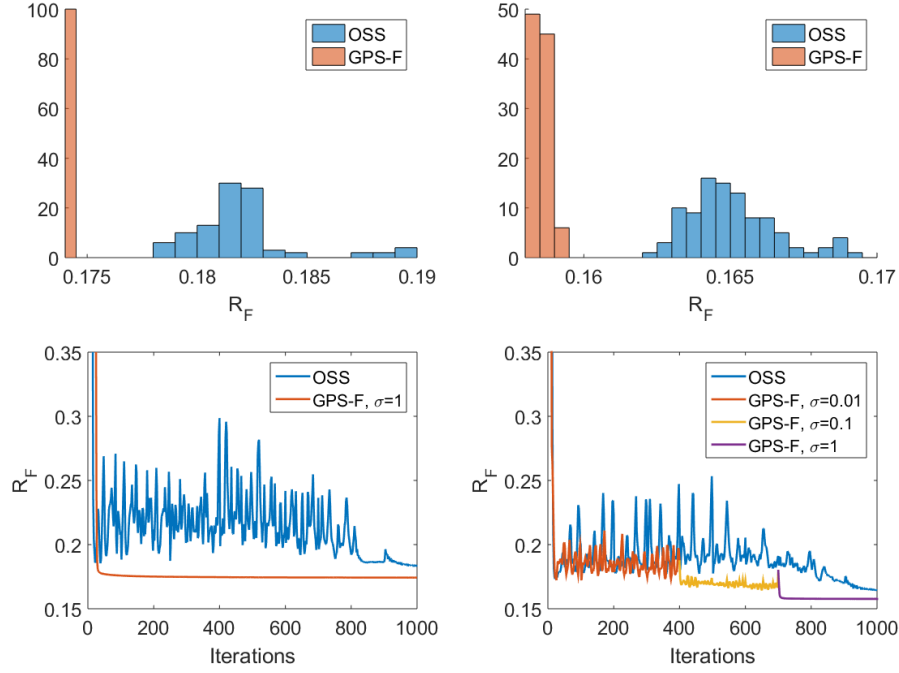


Fig. 6: Histogram (first row) and convergence (second row) of OSS and GPS-F on nanorice1 (first column) and nanorice2 (second column). The results of HIO are omitted due to lack of convergence.

the mobility to skip local minima. We recommend start with small  $\sigma$  and then gradually increase it until the iterative process reaches a stable minimum.

## 5. Conclusion

In this work, we have developed a fast and robust phase retrieval algorithm GPS for the reconstruction of images from noisy diffraction intensities. Similar to [34], the Moreau-Yosida regularization was used to relax the hard constraints considered in the noiseless model. GPS utilizes a primal-dual algorithm and a noise-removal technique, in which the  $\ell_2$  gradient smoothing is effectively applied on either real or Fourier space of the dual variable. GPS shows more reliable and consistent results than OSS, HIO for the reconstruction of weakly scattered objects such as biological specimens. Looking forward, we aim to explore the role of dual variables in non-convex optimization. Smoothing the dual variable, which is equivalent to smoothing the gradient of convex conjugate, represents a new and effective technique that can in principle be applied to other non-smooth, non-convex problems.

## ACKNOWLEDGMENT

We thank Professor Jose A. Rodriguez for help and providing experimental data of the yeast spore. This work was supported by STROBE: A National Science Foundation Science & Technology Center, under Grant No. DMR 1548924 as well as DOE-DE-SC0013838.

## APPENDIX

We cover some mathematics background in this section.

### *Proximal Mapping on complex domain*

We extend the definition of the Moreau-Yosida regularization [32, 33] to complex domain. Let  $G \in \mathbb{C}^{n \times n}$  be a Hermitian positive definite matrix. The Moreau-Yosida regularization of a lower semi-continuous extended-real-valued function  $h : \mathbb{C}^n \rightarrow (-\infty, \infty]$ , associated with  $G$ , is defined by

$$h_G(u) := \inf_{v \in \mathbb{C}^n} \left\{ h(v) + \frac{1}{2} \|v - u\|_{G^{-1}}^2 \right\}. \quad (32)$$

We see that  $h_G$  converges pointwise to  $h$  as  $\|G\| \rightarrow 0^+$ . In the special case where  $G = tI$  is a multiple of identity matrix with  $t > 0$ ,  $h_G$  reduces to

$$h_t(u) := \inf_{v \in \mathbb{C}^n} \left\{ h(v) + \frac{1}{2t} \|v - u\|^2 \right\}. \quad (33)$$

For any characteristic function  $h = \mathcal{I}_X$  of a closed set  $X \subset \mathbb{C}^n$ ,

$$h_t(u) := \frac{1}{2t} \inf_{v \in X} \|v - u\|^2 = \frac{1}{2t} \|u - \text{proj}_X(u)\|^2 \quad (34)$$

is  $\frac{1}{2t}$  of the squared  $\ell_2$  distance from  $u$  to the set  $X$ . Similar idea of relaxing a characteristic function into a distance function has been successfully applied to the quantization problem of deep neural networks in [34]. Taking  $X = \{z \in \mathbb{R}^n : |z| = b\}$  to be the magnitude constraint set and  $\sigma > 0$ , we first relax  $g = \mathcal{I}_{|z|=b}$  in (3) into

$$g_\sigma(z) := \frac{1}{2\sigma} \inf_{|v|=b} \|v - z\|^2. \quad (35)$$

Since  $\text{proj}_{|z|=b}(z) = b \odot \exp(\mathbf{i} \cdot \arg(z))$  is the projection of  $z$  onto the set  $\{z \in \mathbb{R}^n : |z| = b\}$ , a simple calculation shows that

$$g_\sigma(z) = \frac{1}{2\sigma} \|b \odot \exp(\mathbf{i} \cdot \arg(z)) - z\|^2 = \frac{1}{2\sigma} \|(b - |z|) \odot \exp(\mathbf{i} \cdot \arg(z))\|^2 = \frac{1}{2\sigma} \||z| - b\|^2 \quad (36)$$

is a least squares fidelity, which measures the difference between the observed magnitudes and fitted ones.

### *Generalized Legendre-Fenchel transformation*

We can express any function  $h : \mathbb{C}^n \rightarrow \mathbb{R}$  as a function  $\tilde{h}$  defined on  $\mathbb{R}^{2n}$  in the following way

$$h(u) := \tilde{h}(\text{Re}(u), \text{Im}(u)) = \tilde{h}(\tilde{u}), \quad (37)$$

where  $\tilde{u} = [\text{Re}(u); \text{Im}(u)] \in \mathbb{R}^{2n}$  and  $\tilde{h} : \mathbb{R}^{2n} \rightarrow \mathbb{R}$ . We define that  $h$  is convex, if  $\tilde{h}$  is convex on  $\mathbb{R}^{2n}$ . Note that for any  $u, y \in \mathbb{C}^n$ ,

$$\text{Re}\langle u, y \rangle = \langle \text{Re}(u), \text{Re}(y) \rangle + \langle \text{Im}(u), \text{Im}(y) \rangle = \langle \tilde{u}, \tilde{y} \rangle. \quad (38)$$

We propose to generalize the Legendre-Fenchel transformation (a.k.a. convex conjugate) [48] of an extended-real-valued convex function  $h$  defined on  $\mathbb{C}^n$  as

$$h^*(y) := \sup_{u \in \mathbb{C}^n} \left\{ \text{Re}\langle y, u \rangle - h(u) \right\}. \quad (39)$$

### Infimal Convolution & Moreau-Yoshida decomposition

By Eq. (5),  $f_G$  is the infimal convolution [49] between the convex functions  $f$  and  $\frac{1}{2}\|\cdot\|_{G^{-1}}^2$ . The Moreau-Yoshida decomposition uncouples the Legendre-Fenchel transformation of  $f_G$

$$f_G^*(y) = f^*(y) + \left(\frac{1}{2}\|\cdot\|_{G^{-1}}^2\right)^*(y), \quad (40)$$

where  $\left(\frac{1}{2}\|\cdot\|_{G^{-1}}^2\right)^*(y) = \frac{1}{2}\|y\|_G^2$ . Here we remark that  $G$  only needs to be positive semi-definite in Eq. (5), as  $f_G$  can take the extended value  $\infty$ . In this case, although  $G^{-1}$  does not exist, since  $f_G$  is convex and lower semi-continuous, and the strong duality  $f_G^{**} = f_G$  holds here, we can re-define  $f_G$  via the biconjugate as

$$\begin{aligned} f_G(u) = f_G^{**}(u) &= (f_G^*(y))^* = \sup_{y \in \mathbb{C}^n} \left\{ \operatorname{Re}\langle u, y \rangle - f_G^*(y) \right\} = \sup_{y \in \mathbb{C}^n} \left\{ \operatorname{Re}\langle u, y \rangle - f^*(y) - \frac{1}{2}\|y\|_G^2 \right\} \\ &= \sup_{y \in \mathcal{S}^*} \left\{ \operatorname{Re}\langle u, y \rangle - \frac{1}{2}\|y\|_G^2 \right\}. \end{aligned} \quad (41)$$

### References

1. R. Millane, "Phase retrieval in crystallography and optics," J. Opt. Soc. Am. A **7**, 394–411 (1990).
2. J. Miao, T. Ishikawa, I. Robinson, and M. Murnane, "Beyond crystallography: Diffractive imaging using coherent x-ray light sources," Science **348**, 530–535 (2015).
3. Y. Shechtman, Y. Eldar, O. Cohen, H. Chapman, J. Miao, and M. Segev, "Phase retrieval with application to optical imaging: a contemporary overview," IEEE Signal Process. Mag. **32**, 87–109 (2015).
4. S. Marchesini, "Invited article: A unified evaluation of iterative projection algorithms for phase retrieval," Rev. Sci. Instrum. **78**, 011301 (2007).
5. J. Miao, D. Sayre, and H. Chapman, "Phase retrieval from the magnitude of the fourier transforms of nonperiodic objects," J. Opt. Soc. Am. A. **15**, 1662–1669 (1998).
6. R. Gerchberg and W. Saxton, "A practical algorithm for the determination of the phase from image and diffraction plane pictures," Optik **35**, 237–246 (1972).
7. J. Fienup, "Reconstruction of an object from the modulus of its fourier transform," Opt. Lett. **3**, 27–29 (1978).
8. J. Miao, P. Charalambous, J. Kirz, and D. Sayre, "Extending the methodology of x-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens," Nature **400** (1999).
9. H. Chapman and K. Nugent, "Coherent lensless x-ray imaging," Nat. Photon. **4**, 833–839 (2010).
10. I. Robinson and R. Harder, "Coherent x-ray diffraction imaging of strain at the nanoscale," Nat. Mater. **8**, 291 (2009).
11. J. Miao and D. Sayre, "On possible extensions of x-ray crystallography through diffraction-pattern oversampling," Acta Crystallogr. A **56**, 596–605 (2000).
12. J. Fienup, "Phase retrieval algorithms: a comparison," Appl. Opt. **21**, 2758–2769 (1982).
13. C. Chen, J. Miao, C. Wang, and T. Lee, "Application of optimization technique to noncrystalline x-ray diffraction microscopy: Guided hybrid input-output method," Phys. Rev. B **76**, 064113 (2007).
14. K. Raines, S. Salha, R. Sandberg, H. Jiang, J. Rodriguez, B. Fahimian, H. Kapteyn, J. Du, and J. Miao, "Three-dimensional structure determination from a single view," Nature **463** **7278**, 214–7 (2010).
15. 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," Opt. Express **20**, 16650–16661 (2012).
16. J. Rodriguez, R. Xu, C. Chen, Y. Zou, and J. Miao, "Oversampling smoothness (OSS): an effective algorithm for phase retrieval of noisy diffraction intensities," J. Appl. Crystallogr. **46**, 312–318 (2013).
17. H. Bauschke, P. Combettes, and D. Luke, "Hybrid projection-reflection method for phase retrieval," J. Opt. Soc. Am. A. **20** **6**, 1025–34 (2003).
18. J. Douglas and H. Rachford, "On the numerical solution of heat conduction problems in two and three space variables," Trans. Amer. Math. Soc. pp. 421–439 (1956).
19. H. Bauschke, P. Combettes, and D. Luke, "Phase retrieval, error reduction algorithm, and fienup variants: a view from convex optimization," J. Opt. Soc. Am. A. **19** **7**, 1334–45 (2002).
20. V. Elser, "Phase retrieval by iterated projections," J. Opt. Soc. Am. A. **20** **1**, 40–55 (2003).
21. D. Luke, "Relaxed averaged alternating reflections for diffraction imaging," Inv. Prob. pp. 37–50 (2005).

22. P. Chen and A. Fannjiang, "Fourier phase retrieval with a single mask by Douglas-Rachford algorithms," *Appl. Comput. Harmon. Anal.* (2016).
23. E. Candès, X. Li, and M. Soltanolkotabi, "Phase retrieval via wirtinger flow: Theory and algorithms," *IEEE Trans. Info. Theory* **61**, 1985–2007 (2015).
24. H. Chang, Y. Lou, M. Ng, and T. Zeng, "Phase retrieval from incomplete magnitude information via total variation regularization," *SIAM J. Sci. Comput.* **38**, A3672–A3695 (2016).
25. H. Chang, S. Marchesini, Y. Lou, and T. Zeng, "Variational phase retrieval with globally convergent preconditioned proximal algorithm," *SIAM J. Imaging Sci.* **11**, 56–93 (2018).
26. Z. Wen, C. Yang, X. Liu, and S. Marchesini, "Alternating direction methods for classical and ptychographic phase retrieval," *Inv. Prob.* **28** (2012).
27. A. Chai, M. Moscoso, and G. Papanicolaou, "Array imaging using intensity-only measurements," *Inv. Prob.* **27** (2011).
28. E. Candès, Y. Eldar, T. Strohmer, and V. Voroninski, "Phase retrieval via matrix completion," *SIAM J. Imaging Sci.* **6**, 199–225 (2013).
29. E. Candès, T. Strohmer, and V. Voroninski, "Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming," *Comm. Pure Appl. Math.* **66**, 1241–1274 (2013).
30. I. Waldspurger, A. d'Aspremont, and S. Mallat, "Phase recovery, maxcut and complex semidefinite programming," *Math. Program.* **149**, 47–81 (2015).
31. P. Yin and J. Xin, "Phaseliftoff: an accurate and stable phase retrieval method based on difference of trace and frobenius norms," *Commun. Math. Sci.* **13**, 1033–1049 (2015).
32. J.-J. Moreau, "Proximité et dualité dans un espace hilbertien," *Bull. de la Société Mathématique de France* **93**, 273–299 (1965).
33. K. Yosida, *Functional analysis* (Springer, Berlin, 1964).
34. P. Yin, S. Zhang, J. Lyu, S. Osher, Y. Qi, and J. Xin, "Binaryrelax: A relaxation approach for training deep neural networks with quantized weights," *UCLA CAM Rep.* 18-05; *SIAM J. Imaging Sci.*, to appear (2018).
35. R. Glowinski and A. Marroco, "Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de dirichlet non linéaires," *Revue française d'automatique, informatique, recherche opérationnelle. Analyse numérique* **9**, 41–76 (1975).
36. S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations Trends Mach. Learn.* **3**, 1–122 (2011).
37. M. Yan and W. Yin, "Self equivalence of the alternating direction method of multipliers," in *Splitting Methods in Communication, Imaging, Science, and Engineering*, (Springer, 2016), pp. 165–194.
38. A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imaging Vis.* **40**, 120–145 (2010).
39. E. Esser, X. Zhang, and T. Chan, "A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science," *SIAM J. Imaging Sci.* **3**, 1015–1046 (2010).
40. T. Goldstein, M. Li, X. Yuan, E. Esser, and R. Baraniuk, "Adaptive primal-dual hybrid gradient methods for saddle-point problems," *arXiv preprint arXiv:1305.0546* (2013).
41. D. O'Connor and L. Vandenberghe, "On the equivalence of the primal-dual hybrid gradient method and douglas-rachford splitting," *optimization-online.org* (2017).
42. J. Miao, Y. Nishino, Y. Kohmura, B. Johnson, C. Song, S. H. Risbud, and T. Ishikawa, "Quantitative image reconstruction of gan quantum dots from oversampled diffraction intensities alone," *Phys. Rev. Lett.* **95** **8**, 085503 (2005).
43. J. Miao, K. Hodgson, T. Ishikawa, C. Larabell, M. Legros, and Y. Nishino, "Imaging whole escherichia coli bacteria by using single-particle x-ray diffraction," *Proc. Natl. Acad. Sci. USA* **100**, 110–112 (2003).
44. C. Song, K. Tono, J. Park, T. Ebisu, S. Kim, H. Shimada, S. Kim, M. Gallagher-Jones, D. Nam, T. Sato, T. Togashi, K. Ogawa, Y. Joti, T. Kameshima, S. Ono, T. Hatsui, S. Iwata, M. Yabashi, and T. Ishikawa, "Multiple application X-ray imaging chamber for single-shot diffraction experiments with femtosecond X-ray laser pulses," *J. Appl. Crystallogr.* **47**, 188–197 (2014).
45. H. Jiang, C. Song, C.-C. Chen, R. Xu, K. S. Raines, B. P. Fahimian, C.-H. Lu, T.-K. Lee, A. Nakashima, J. Urano, T. Ishikawa, F. Tamanoi, and J. Miao, "Quantitative 3d imaging of whole, unstained cells by using x-ray diffraction microscopy," *Proc. Natl. Acad. Sci. USA* **107**, 11234–11239 (2010).
46. S. Kassemeyer, J. Steinbrener, L. Lomb, E. Hartmann, A. Aquila, A. Barty, A. V. Martin, C. Y. Hampton, S. Bajt, M. Barthelmeß, T. R. Barends, C. Bostedt, M. Bott, J. D. Bozek, N. Coppola, M. Cryle, D. P. DePonte, R. B. Doak, S. W. Epp, B. Erk, H. Fleckenstein, L. Foucar, H. Graafsma, L. Gumprecht, A. Hartmann, R. Hartmann, G. Hauser, H. Hirsemann, A. Hömke, P. Holl, O. Jönsson, N. Kimmel, F. Krasniqi, M. Liang, F. R. Maia, S. Marchesini, K. Nass, C. Reich, D. Rolles, B. Rudek, A. Rudenko, C. Schmidt, J. Schulz, R. L. Shoeman, R. G. Sierra, H. Soltau, J. C. H. Spence, D. Starodub, F. Stellato, S. Stern, G. Stier, M. Svenda, G. Weidenspointner, U. Weierstall, T. A. White, C. Wunderer, M. Frank, H. N. Chapman, J. Ullrich, L. Strüder, M. J. Bogan, and I. Schlichting, "Femtosecond free-electron laser x-ray diffraction data sets for algorithm development," *Opt. Express* **20**, 4149–4158 (2012).
47. F. Maia, "The Coherent X-ray Imaging Data Bank," *Nat. Methods* **9**, 854 (2012).
48. R. T. Rockafellar, *Convex analysis* (Princeton university press, 2015).
49. R. Phelps, *Convex Functions, Monotone Operators and Differentiability* (Springer, 1991).