Accelerated Wirtinger Flow With Score-Based Image Priors for Holographic Phase Retrieval in Poisson-Gaussian Noise Conditions

Zongyu Li, *Member, IEEE*, Jason Hu, *Student Member, IEEE*, Xiaojian Xu, *Member, IEEE*, Liyue Shen, *Member, IEEE*, and Jeffrey A. Fessler, *Fellow, IEEE*

***Abstract*—Phase retrieval (PR) is a crucial problem in many imaging applications. This study focuses on holographic phase retrieval in situations where the measurements are degraded by a combination of Poisson and Gaussian noise, as commonly occurs in optical imaging systems. We propose a new algorithm called "AWFS" that uses accelerated Wirtinger flow (AWF) with a learned score function as a generative prior. Specifically, we formulate the PR problem as an optimization problem that incorporates both data fidelity and regularization terms. We calculate the gradient of the log-likelihood function for PR and determine its corresponding Lipschitz constant. Additionally, we introduce a generative prior in our regularization framework by using score matching to capture information about the gradient of image prior distributions. We provide theoretical analysis that establishes a critical-point convergence guarantee for one version of the proposed algorithm. The results of our simulation experiments on three different datasets show the following. 1) By using the PG likelihood model, a practical version of the proposed algorithm improves reconstruction compared to algorithms based solely on Gaussian or Poisson likelihoods. 2) The proposed score-based image prior method leads to better reconstruction quality than a method based on denoising diffusion probabilistic model (DDPM), as well as a plug-and-play alternating direction method of multipliers (PnP-ADMM) and regularization by denoising (RED).**

***Index Terms*—Phase retrieval, poisson-Gaussian noise, score-based diffusion models, wirtinger flow.**

I. Introduction

PHASE retrieval (PR) is a nonlinear inverse problem, where the goal is to recover a signal from the (square of) magnitude-only measurements that are corrupted by noise [2]. This problem has applications in astronomy [3], X-ray crystallography [4], optical imaging [5], Fourier ptychography [6], [7], [8], [9] and coherent diffractive imaging (CDI) [10]. For example, in holographic CDI, a coherent beam source illuminates a sample of interest and a reference. When the beam hits the sample, it generates secondary electromagnetic waves that propagate until they reach a detector. By measuring the photon flux, the detector can capture and record a diffraction pattern. This pattern is roughly proportional to the square of Fourier transform magnitude of electric field associated with the illuminated objects [11], [12]. Recovering the structure of the sample from the diffraction pattern generated by the mix of sample and reference is a nonlinear inverse problem known as holographic PR. One approach to this problem is maximum a posteriori (MAP) estimation:

$$\begin{aligned}\hat{\boldsymbol{x}} &= \arg\max_{\boldsymbol{x}\in\mathbb{R}^N} p(\boldsymbol{x}|\boldsymbol{y},\bar{\boldsymbol{b}},\boldsymbol{A},\boldsymbol{r}) \\ &= \arg\min_{\boldsymbol{x}\in\mathbb{R}^N} g(\boldsymbol{x};\boldsymbol{A},\boldsymbol{y},\bar{\boldsymbol{b}},\boldsymbol{r}) + h(\boldsymbol{x}), \end{aligned} \tag{1}$$

where $\boldsymbol{x}$ denotes a real latent image to recover, $\boldsymbol{y}$ is the recorded measurement vector, $\bar{\boldsymbol{b}}$ denotes the mean of background measurements, and $\boldsymbol{A} \in \mathbb{C}^{M\times N}$ denotes the system matrix in holographic PR, where M denotes the number of measurements and N denotes the dimension of $\boldsymbol{x}$. The known reference image $\boldsymbol{r}$ provides additional information to reduce the ambiguity of $\hat{\boldsymbol{x}}$; using an extended reference is a common technique in holographic CDI [13], [14]. Following Bayes' rule, we denote $g(\boldsymbol{x}) = -\log p(\boldsymbol{y},\boldsymbol{A},\boldsymbol{r}|\boldsymbol{x})$ and $h(\boldsymbol{x}) = -\log p(\boldsymbol{x})$ as the data fidelity term and the regularization term, respectively. For simplicity we assume $\boldsymbol{x}$ is real, so $\nabla g(\boldsymbol{x})$ denotes the real component of the gradient of log-likelihood. Hence, all priors are trained with real-valued datasets and are applied to the real components of $\boldsymbol{x}$ as well. The method can be extended to complex images.

In practice, often the measurements $\boldsymbol{y}$ are contaminated by both Poisson and Gaussian (PG) noise. The Poisson distribution is due to the light photons and dark current [15]. The Gaussian statistics stem from Johnson-Nyquist noise (or thermal noise) in the sensor electronics [16], [17], [18]. The sum of both leads to measurements with PG noise. Fig. 1 illustrates the PG noise statistics in the holographic PR. Because the PG likelihood in (4) is complicated, most previous works [5], [6], [8], [19], [20], [21], [22], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31],

Received 5 October 2023; revised 20 April 2024 and 11 July 2024; accepted 2 September 2024. Date of current version 27 September 2024. This work was supported in part by LANL grant, and in part by KLA gift. An earlier version of this paper was presented in part at the 2023 NeurIPS Workshop on Deep Learning and Inverse Problems [DOI: 10.48550/arXiv.2305.07712]. The associate editor coordinating the review of this article and approving it for publication was Prof. Bo Zhao. *(Zongyu Li and Jason Hu contributed equally to this work.) (Corresponding author: Zongyu Li.)*

The authors are with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109-2122 USA (e-mail: zonyul@umich.edu; jashu@umich.edu; xjxu@umich.edu; liyues@umich.edu; fessler@umich.edu).

Code for reproducing the results is available at https://github.com/ZongyuLi-umich/2023-PGPR.

This article has supplementary downloadable material available at https://doi.org/10.1109/TCI.2024.3458418, provided by the authors.

Digital Object Identifier 10.1109/TCI.2024.3458418

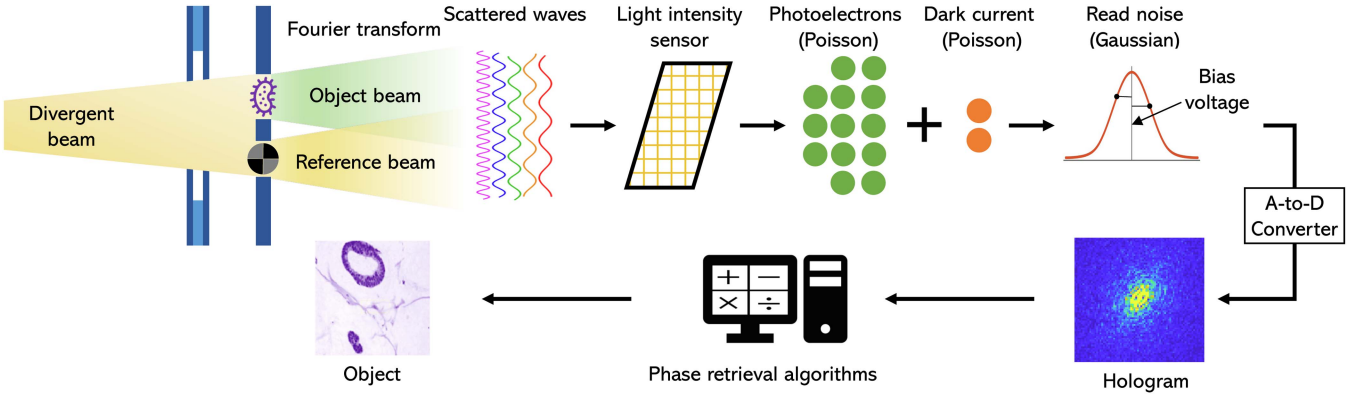


Fig. 1. Illustration of Poisson and Gaussian noise statistics in Fourier transform holographic phase retrieval.

[32], [33], [34], [35], [36], [37], [38], [39], [40], [41], [42], [43], [44], [45], [46] approximate the Poisson noise statistics by the central limit theorem and work with a Gaussian approximation or use the Poisson maximum likelihood model but simply disregard Gaussian readout noise. Other more complicated approximation methods have also been proposed, such as the shifted Poisson model [47], the unbiased inverse transformation of a generalized Anscombe transform [9], [48], and a majorize-minimize algorithm [49]. However, these approximate methods can provide suboptimal solutions. Apart from the likelihood modeling, the regularizer $h(\boldsymbol{x})$ provides prior information about underlying object characteristics that may aid in resolving ill-posed inverse problems. Simple choices of $h(\boldsymbol{x})$ such as total variation (TV) and higher order generalizations including total generalized variation (TGV) [50], [51] and the L1-norm of coefficients of wavelet transform [52] have been studied extensively. More recently, deep learning (DL)-integrated algorithms for solving inverse problems in computational imaging have been reported to be the state-of-the-art [53]. A trained network can serve as an object prior for regularizing the reconstructed image so it remains near a learned manifold [54]. Incorporating a trained denoising network as a regularizer $h(\cdot)$ led to methods such as plug-and-play (PnP) [55], [56], [57] and regularization by denoising (RED) [58]. In contrast to training a denoiser using clean images, there is growing interest in self-supervised image denoising approaches that do not require clean data as the training target [59], [60], [61].

In addition to training a denoiser as regularizer, generative model-based priors have also been proposed [62], [63]. Recently, diffusion models have gained significant traction for image generation [64], [65], [66], [67]. These probabilistic image generation models start with a clean image and gradually increase the level of noise added to the image, resulting in white Gaussian noise. Then in the reverse process, a neural network is trained to learn the noise in each step to generate or sample a clean image as in the original data distribution. The score-based diffusion models estimate the gradients of data distribution and can be used as plug-and-play priors for inverse problems [68] such as image deblurring and MRI and CT reconstruction [69], [70], [71], [72], [73], [74]. However, the realm of using score-based models to perform phase retrieval is relatively unexplored; previous relevant works [68], [75] applied denoising diffusion probabilistic modeling (DDPM) to PR but with less realistic system models and under solely Gaussian or Poisson noise statistics.

In summary, the remainder of the paper is organized as follows:

- Section II provides an overview of forward and noise models used for phase retrieval and deep learning regularizer methods that have been used to solve the PR problem.
- Section III presents our proposed algorithm known as accelerated Wirtinger flow with a score-based image prior (i.e., $\nabla h(\boldsymbol{x})$ in (1)) to address the challenge of holographic phase retrieval (PR) problem in the presence of Poisson and Gaussian (PG) noise statistics. Theoretically, we derive a Lipschitz constant for the holographic PR's PG log-likelihood and provide a guarantee that sequences generated by the proposed algorithm converge to critical points of the cost function.
- Section IV consists of simulation experiments which demonstrate that: 1) Algorithms using the PG likelihood model yield superior reconstructions compared to those relying solely on either the Poisson or Gaussian likelihood models. 2) With the proposed score-based prior as regularization, the proposed approach generates higher quality reconstructions and is more robust to variations of noise levels (without any parameter tuning) than alternative state-of-the-art methods.
- Section IV and Section V consist of a discussion and conclusion about the results, comparisons with existing approaches, and limitations of our approach along with future work.

II. Background and Related Work

This section reviews technical background for phase retrieval (PR) and summarizes previous works and their limitations.

A. Forward Models for PR

Many previous works [19], [20], [21], [22], [23], [27], [28], [29], [31], [32], [34], [45], [49], [75], [76] modeled the system matrix $\boldsymbol{A}$ in (1) as i.i.d. random Gaussian or randomly masked Fourier transform; these assumptions simplify the PR

problem and lead to elegant mathematical derivations (e.g., spectral initialization [33], [77]), but they are less related to optical imaging systems used in practical PR [78]. Practical PR involves canonical Fourier transform-based system matrices such as Fresnel (near-field) PR [79], [80], holographic PR [81], ptychographic PR [82], [83] and Fraunhofer (far-field) PR [78], [84]. The canonical Fourier transform-based PR probem is more difficult as it introduces more ambiguities, such as translation and conjugate flipping. Some previous works also remove the square of the Fourier transform magnitude (see (3)) [32], [34], [45], [75]. However, this square of magnitude indicates the amount of wavelength-weighted power emitted by a light source per unit area, so its removal reduces the physics realism of a model.

B. Noise Models for PR

Gaussian PR: For methods that assume the elements of $\boldsymbol{y}$ follow independent Gaussian distributions $\boldsymbol{y} \sim \mathcal{N}(|\boldsymbol{A}\boldsymbol{x}|^2 + \bar{\boldsymbol{b}}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}$, the data fidelity term $g(\boldsymbol{x})$ in (1) becomes $g_{\text{Gau}}(\boldsymbol{x}) \triangleq \|\boldsymbol{y} - \bar{\boldsymbol{b}} - |\boldsymbol{A}\boldsymbol{x}|^2\|_2^2$. To solve the corresponding MAP optimization problem, a popular method is Wirtinger flow (WF) [21], [22], [23], [33] using the Wirtinger gradient: $\nabla g_{\text{Gau}}(\boldsymbol{x}) = 4\boldsymbol{A}' \operatorname{diag}\{|\boldsymbol{A}\boldsymbol{x}|^2 - \boldsymbol{y} + \bar{\boldsymbol{b}}\}\boldsymbol{A}\boldsymbol{x}$. To determine an appropriate step size for the Wirtinger gradient, one can use its Lipschitz constant or methods such as empirical trial and error, backtracking line search, or observed Fisher information [42]. To further accelerate WF, one can use Nesterov's momentum methods [85] or optimized gradient methods [86], leading to the accelerated Wirtinger flow (AWF) [8], [87], [88], [89] that is commonly used in solving PR problems. Apart from WF, other methods such as matrix-lifting [5], [19], [20], error reduction (ER) [25], hybrid input-output (HIO) [90], majorize-minimize (MM) [24] and alternating direction method of multipliers (ADMM) [30] have also been proposed.

However, in the Holographic PR case, due to the Fourier transform, the components of $|\boldsymbol{A}\boldsymbol{x}|^2$ vary greatly in magnitude, and in particular, the low frequency components associated with the DC term of $\boldsymbol{x}$ is much larger than the other components, making this a poor approximation. Algorithms based on such assumption will lead to a low frequency-only reconstruction (see Supplement). Instead, when we initialize the algorithm using the reconstruction from the spectral initialization $\boldsymbol{x}_0$, we may estimate $\boldsymbol{\Sigma} = \operatorname{Diag}(\sqrt{|\boldsymbol{A}\boldsymbol{x}_0|^2 + \bar{\boldsymbol{b}}})$, where the square root is taken element-wise. Using this approach, the log-likelihood of the Gaussian only model can be computed readily and applied in conjunction with the accelerated score based method.

Following [91], [92], we also took the square root of the measurement and then assumed the resulting noise model to be Gaussian to compute the log likelihood as a comparison method. However, the results show that the reconstructed images using the amplitude loss function are much noisier than the other three methods and the quantitative metrics are also worse. This is due to the mismatch between the noise model used in reconstruction and the measurements noise simulation that is otherwise based on the intensity model. Another inconsistency arises when we need to compute the square root, as we have to truncate any negative measurements resulting from the Gaussian noise. Section IV shows a comparison between these models. **Poisson PR.** The Poisson ML model assumes $\boldsymbol{y} \sim \text{Poisson}(|\boldsymbol{A}\boldsymbol{x}|^2 + \bar{\boldsymbol{b}})$, so that $g(\boldsymbol{x})$ in (1) has the form: $g_{\text{Pois}}(\boldsymbol{x}) \triangleq \boldsymbol{1}'(|\boldsymbol{A}\boldsymbol{x}|^2 + \bar{\boldsymbol{b}}) - \boldsymbol{y}' \log(|\boldsymbol{A}\boldsymbol{x}|^2 + \bar{\boldsymbol{b}})$. Similar to the Gaussian case, one can also apply WF [76] with $\nabla g_{\text{Pois}}(\boldsymbol{x}) = 2\boldsymbol{A}\boldsymbol{x} \odot (\boldsymbol{1} - \boldsymbol{y} \oslash (|\boldsymbol{A}\boldsymbol{x}|^2 + \bar{\boldsymbol{b}}))$, where $\odot$ and $\oslash$ denote element-wise multiplication and division, respectively.

Algorithm 1: Phase Retrieval Via Wirtinger Flow.

Require: Measurement $\boldsymbol{y}$, system matrix $\boldsymbol{A}$, initialization of image $\boldsymbol{x}_0$, regularizer $h(\cdot)$.
  for $k = 1 : K$ **do**
    if Gaussian noise model is used **then**
      Compute $\nabla g_{\text{Gau}}(\boldsymbol{x}_k) = 4\boldsymbol{A}' \operatorname{diag}\{|\boldsymbol{A}\boldsymbol{x}_k|^2 - \boldsymbol{y} + \bar{\boldsymbol{b}}\} \boldsymbol{A}\boldsymbol{x}_k$.
    else if Poisson noise model is used **then**
      Compute $\nabla g_{\text{Pois}}(\boldsymbol{x}_k) = 2\boldsymbol{A}\boldsymbol{x}_k \odot (\boldsymbol{1} - \boldsymbol{y} \oslash (|\boldsymbol{A}\boldsymbol{x}_k|^2 + \bar{\boldsymbol{b}}))$.
    end if
    Compute gradient of the regularizer $\nabla h(\boldsymbol{x}_k)$.
    Compute step size μ_k.
    Set $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \mu_k(\nabla g(\boldsymbol{x}_k) + \nabla h(\boldsymbol{x}_k))$.
  end for
Return $\boldsymbol{x}_K$

Algorithm 1 summarizes the conventional WF approaches for PR. The Gaussian and Poisson models are both suboptimal for practical scenarios where the measurements are corrupted by PG noise.

C. Regularizers for PR

Non-DL Methods: Traditionally, total variation (TV) has been used a regularizer in PR [46]. This regularizer has a denoising effect on the image while preserving sharp boundaries. Similarly, wavelet based methods [52] transform the image to the wavelet domain where it becomes simpler to denoise the image without distorting its main features. More recently, plug and play methods involve alternating between using a denoiser to denoise the image and a data fidelity update step to ensure that the image conforms with the measurement. Block matching 3D (BM3D) is a fast and effective general purpose denoising algorithm, so its use in PR has been studied [93].

PnP and RED: Both PnP and RED are widely adopted in a variety of inverse problems [94], [95], [96], [97], [98], [99], [100]. For example, Wei et al. developed a policy network to automatically search for the best tuning parameters in PnP frameworks, which can be effectively learned with mixed model-free and model-based deep reinforcement learning [98]. By implicitly representing the prior $h(\cdot)$ in (1) by an image denoiser, plug-and-play (PnP) methods [101], [102] were proposed to allow the integration of physical measurement models and powerful DL-based denoisers as image priors [57]. The model-agnostic nature of the denoiser allows PnP methods to be applied to multiple imaging problems using a single DL denoiser by changing only the imaging model. Regularization by denoising (RED) [103],

[104] is an algorithm closely related to PnP that uses a denoising engine in defining the regularization of the inverse problems. For PR problems, many PnP methods were proposed [91], [92], [93], [105], [106]. For instance, Metzler et al. proposed to leverage the regularization-by-denoising framework and a convolutional neural network as a PnP denoiser [91]; Wang et al. combined iterative methods from PR with RED, pointing to the direction of hybrid methods that integrate learned PnP priors in conventional algorithms. The effectiveness of PnP inspired this study of using a generative PnP image prior using score matching to address the PR problem.

Score Function and Diffusion Models: Let $p_{\boldsymbol{\theta}}(\boldsymbol{x})$ denote a model for the prior distribution of the latent image $\boldsymbol{x}$; the score function is then defined as[1] $s_{\boldsymbol{\theta}}(\boldsymbol{x}) = \nabla_{\boldsymbol{x}} \log p_{\boldsymbol{\theta}}(\boldsymbol{x})$. Consider a sequence of positive noise scales (for white Gaussian $\mathcal{N}(0, \sigma_k^2)$): $\sigma_1 > \sigma_2 > \cdots > \sigma_K$, with σ_K being small enough so that noise of this level does not visibly affect the image, and σ_1 depending on the application. Score matching can be used to train a noise conditional score network (NCSN) [64], [107] as follows:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{k=1}^{K} \mathbb{E}_{\boldsymbol{x},\tilde{\boldsymbol{x}}} \left[\left(s_{\boldsymbol{\theta}}(\boldsymbol{x}, \sigma_k) - \frac{\boldsymbol{x} - \tilde{\boldsymbol{x}}}{\sigma_k^2} \right)^2 \right],$$
$$\text{where} \quad \boldsymbol{x} \sim p(\boldsymbol{x}), \quad \tilde{\boldsymbol{x}} \sim \boldsymbol{x} + \mathcal{N}(0, \sigma_k^2 \boldsymbol{I}). \tag{2}$$

With enough data, the neural network $s_{\boldsymbol{\theta}}(\boldsymbol{x}, \sigma)$ is expected to learn the distribution $p_\sigma(\boldsymbol{x}) = \int p(\boldsymbol{x}) p_\sigma(\tilde{\boldsymbol{x}}|\boldsymbol{x}) d\boldsymbol{x}$ where $p_\sigma(\tilde{\boldsymbol{x}}|\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}, \sigma^2 \boldsymbol{I})$. To sample from the prior, the method of Langevin dynamics is often used [64].

To leverage diffusion models for solving inverse problems, previous methods generally recast the reconstruction problem as a conditional generation or sampling problem [67], [68], [74], [75], [108], [109]. This involves relying on the capacity of diffusion models to produce high-quality images while complying with data-fidelity constraints. However, in applications where data collection is costly, i.e., with a limited amount of training data, it is often challenging to train a diffusion model that can generate high-quality images even in an unconditional way. Under these conditions, we found that the score function learned during training diffusion models can serve as an effective image prior (as demonstrated in Section IV), that can capture certain data characteristics when trained for the denoising prediction in the reverse process of the diffusion model. Similar to previous works [68] that used the score function as a PnP prior, here we also incorporate the score function as a regularization in the optimization objective for solving the PR problem. We believe this is a more efficient scheme for incorporating diffusion priors, especially for applications with a limited amount of training data. Most closely related to this work, [75] applied the DDPM method [65] to a PR problem with Poisson noise. Ref. [67] shows that, under the general framework of score-based generative model with stochastic differential equations, using the score matching method (which is the backbone of our AWFS method) to perform unconditional sampling has the same training objective as the DDPM method. For solving inverse problems, reconstructing the image can be interpreted as maximizing the posterior probability. Under this interpretation, the score matching approach can then be used in conjunction with the proposed acceleration method to perform accelerated gradient descent, greatly reducing the number of steps needed in the WF algorithm. However, the DDPM method does not allow a similar method of acceleration, as its sampling process involves computing linear combinations of the current noisy image and the noise estimate from the neural network.

Thus, to address the phase retrieval challenge with PG noise using the DDPM method, one would have to either: decrease the number of sampling steps without any other alterations to the algorithm, resulting in poorer image quality; or substitute the PG likelihood with a simpler but less precise Poisson or Gaussian likelihood. With score-based models, we can integrate Nesterov's acceleration to enhance algorithm efficiency while maintaining the use of the PG likelihood.

D. Limitations of Previous Works

One main motivation of this work originates from limitations from previous works. Many of them used the unregularized Gaussian, Poisson, and PG methods. For these methods, the objective function simply seeks to maximize the measurement log likelihood. When the measurements are very noisy, the optimal value of the objective function will be attained by a very noisy image. Without regularization, a method cannot enforce any expected image properties. Adding regularizers such as total variation can reduce the level of noise present in the reconstructed image, while not heavily blurring sharp boundaries, but is suboptimal to use when the latent image is not piece-wise uniform. Higher order TV and total generalized variation (TGV) can help to mitigate this problem, but they have higher computation complexity and there is still a tradeoff in the reconstructed image between sharp boundaries and overall smoothness that is unavoidable when not using data driven methods. Instead, DL-based plug-and-play methods are popular to handle the limitations of hand-crafted regularizers, but the noise level of the image will change throughout the iterative process and is unknown. In contrast, the method of training diffusion models to learn the score function of the underlying image distributions (independent of the sensor) is less sensitive to the noise levels and use it as a generative prior in the iterative reconstruction process is relatively unexplored for PR.

III. Proposed Methods

A. Poisson-Gaussian Log-Likelihood for PR

Based on the physical model as demonstrated in Fig. 1, we model the system matrix $\boldsymbol{A}$ by the (oversampled and scaled) discrete Fourier transform applied to a concatenation of the sample $\boldsymbol{x}$, a blank image (representing the holographic separation condition [39]) and a known reference image $\boldsymbol{r}$, so that $\boldsymbol{y}$ follows the Poisson plus Gaussian distribution:

$$\boldsymbol{y} \sim \mathcal{N}(\text{Poisson}\left(|\boldsymbol{A}(\boldsymbol{x})|^2 + \bar{\boldsymbol{b}}\right), \sigma^2 \boldsymbol{I}),$$
$$\boldsymbol{A}(\boldsymbol{x}) \triangleq \alpha \mathcal{F}\{[\boldsymbol{x}, \boldsymbol{0}, \boldsymbol{r}]\}. \tag{3}$$

[1]This definition differs from the score function in statistics where the gradient is taken w.r.t. $\boldsymbol{\theta}$ of $\log p_{\boldsymbol{\theta}}(\boldsymbol{x})$.

Here σ^2 denotes the variance of Gaussian noise, and α denotes a scaling factor (quantum efficiency, conversion gain, etc.) after applying the Fourier transform. Plugging the negative log-likelihood of (3) into (1) leads to

$$g_{\text{PG}}(\boldsymbol{x}) = \sum_{i=1}^{M} g_i(\boldsymbol{x}),$$

$$g_i(\boldsymbol{x}) \triangleq -\log\left(\sum_{n=0}^{\infty} \frac{\mathrm{e}^{-\left(|\boldsymbol{a}_i'\boldsymbol{x}|^2+\bar{b}_i\right)} \cdot \left(|\boldsymbol{a}_i'\boldsymbol{x}|^2+\bar{b}_i\right)^n}{n!} \cdot \frac{\mathrm{e}^{-\left(\frac{(y_i-n)^2}{\sqrt{2}\sigma}\right)}}{\sqrt{2\pi\sigma^2}}\right). \tag{4}$$

Here $\boldsymbol{a}_i'$ denotes the ith row of $\boldsymbol{A}$ (since $\boldsymbol{A}$ is linear). We opt to use WF for estimating $\boldsymbol{x}$ because it is commonly used in practice due to its simplicity and efficiency [33]. The WF algorithm is based on the gradient of (4):

$$\nabla g_{\text{PG}}(\boldsymbol{x}) = 2\boldsymbol{A}' \operatorname{diag}\{\phi_i(|\boldsymbol{a}_i'\boldsymbol{x}|^2+; y_i)\}\boldsymbol{A}\boldsymbol{x},$$

$$\phi(u;v) \triangleq 1 - \frac{s(u,v-1)}{s(u,v)}, \quad s(a,b) \triangleq \sum_{n=0}^{\infty} \frac{a^n}{n!}\mathrm{e}^{-\left(\frac{b-n}{\sqrt{2}\sigma}\right)^2}. \tag{5}$$

Lemma 1: The function $\phi(u)$ is Lipschitz differentiable and the Lipschitz constant for its derivative $\dot{\phi}(u)$ is:

$$\max\{|\ddot{\phi}(u)|\} \triangleq \mu = \left(1-\mathrm{e}^{-1/\sigma^2}\right)\mathrm{e}^{\frac{2y_{\max}-1}{\sigma^2}},$$

$$\text{where} \quad y_{\max} \triangleq \max_{i\in\{1,\ldots,M\}} y_i. \tag{6}$$

The proof of Lemma 1 is given in [110].

To facilitate choosing step sizes, the following Theorem provides a Lipschitz constant for the gradient of the log-likelihood (4).

Theorem 1: Assume $|x_j|$ is bounded above by C for each j, a Lipschitz constant of $\nabla g_{\text{PG}}(\boldsymbol{x})$ is

$$\mathcal{L}(\nabla g_{\text{PG}}) \triangleq 4C^2\|\boldsymbol{A}\|_2^2\|\boldsymbol{A}\|_\infty^2\left(1-\mathrm{e}^{-\frac{1}{\sigma^2}}\right)\mathrm{e}^{\frac{2y_{\max}-1}{\sigma^2}} + 2\|\boldsymbol{A}\|_2^2\left|1-C^2\|\boldsymbol{A}\|_\infty^2\left(1-\mathrm{e}^{-\frac{1}{\sigma^2}}\right)\mathrm{e}^{\frac{2y_{\max}-1}{\sigma^2}}\right|, \tag{7}$$

where $y_{\max} \triangleq \|\boldsymbol{y}\|_\infty$.

Proof: Let $g_{\text{PG}}(\boldsymbol{x})$ denote a function that maps a vector $\boldsymbol{x} \in \mathbb{R}^N$ to a scalar; it is the sum of each $g_i(\boldsymbol{x}) \triangleq \phi_i(|\boldsymbol{a}_i'\boldsymbol{x}|^2+; y_i)$ over $i = 1, \ldots, M$. Let $\boldsymbol{g}(\boldsymbol{x})$ denote a function that maps a vector $\boldsymbol{x} \in \mathbb{R}^N$ to the measurement space $\mathbb{R}^M$; it is the concatenation of each $g_i(\boldsymbol{x})$. So $\nabla g_{\text{PG}}(\boldsymbol{x}) \in \mathbb{R}^N$, $\nabla^2 g_{\text{PG}}(\boldsymbol{x}) \in \mathbb{R}^{N\times N}$, and $\nabla \boldsymbol{g}(\boldsymbol{x}) \in \mathbb{R}^{M\times N}$.

By the chain rule, the Hessian of g_{PG} is

$$\nabla^2 g_{\text{PG}}(\boldsymbol{x}) = 2\boldsymbol{A}'\left(\operatorname{diag}\{\boldsymbol{A}\boldsymbol{x}\}\nabla\boldsymbol{g}(\boldsymbol{x}) + \operatorname{diag}\{\boldsymbol{g}(\boldsymbol{x})\}\boldsymbol{A}\right). \tag{8}$$

Assume $|x_j| \le C$ for each j. Then it follows that $\|\operatorname{diag}\{\boldsymbol{A}\boldsymbol{x}\}\|_2 \le C\|\boldsymbol{A}\|_\infty$ by the construction of matrix-vector multiplication, leading to a Lipschitz constant for $\nabla g_{\text{PG}}(\boldsymbol{x})$:

$$\mathcal{L}(\nabla g_{\text{PG}}) = 2C\|\boldsymbol{A}\|_2\|\boldsymbol{A}\|_\infty\|\nabla\boldsymbol{g}(\boldsymbol{x})\|_2 + 2\|\boldsymbol{A}\|_2^2\|\operatorname{diag}\{\boldsymbol{g}(\boldsymbol{x})\}\|_2. \tag{9}$$

Here $\mathcal{L}(\nabla g_{\text{PG}})$ denotes a Lipschitz constant for ∇g_{PG}, not necessarily the best one. To compute $\|\nabla\boldsymbol{g}(\boldsymbol{x})\|_2$, we substitute the Lipschitz constant of $\dot{\phi}(u)$ into (5) and apply Lemma 1, leading to

$$\|\nabla\boldsymbol{g}(\boldsymbol{x})\|_2 \le 2C\|\boldsymbol{A}\|_2\|\boldsymbol{A}\|_\infty\left(1-\mathrm{e}^{-\frac{1}{\sigma^2}}\right)\mathrm{e}^{\frac{2y_{\max}-1}{\sigma^2}}. \tag{10}$$

To compute $\|\operatorname{diag}\{\boldsymbol{g}(\boldsymbol{x})\}\|_2$, let

$$t \in [b, \max_i\{|\boldsymbol{a}_i'\boldsymbol{x}|^2\}+b] \subseteq \mathcal{T} \triangleq [b, C^2\|\boldsymbol{A}\|_\infty^2+b]. \tag{11}$$

From the fact that $\dot{\phi}(t) \le 1$ by its construction, one can show:

$$\|\operatorname{diag}\{\boldsymbol{g}(\boldsymbol{x})\}\|_2 = \|\boldsymbol{g}(\boldsymbol{x})\|_\infty \le \max_{t\in\mathcal{T}}\{|\dot{\phi}(t)|\} \le \left|1-C^2\|\boldsymbol{A}\|_\infty^2\max\{|\ddot{\phi}(t)|\}\right|. \tag{12}$$

Combining (9), (10) and (12) completes the proof of Theorem 1. □

Theorem 1 extends [110] by considering a a nonlinear transformation ($\boldsymbol{A}\boldsymbol{x} \to |\boldsymbol{A}\boldsymbol{x}|^2$) and a different system matrix $\boldsymbol{A}$. In the practical implementation we approximate (4), with a finite sum following [110]:

$$s(a,b) \approx \sum_{n=0}^{n^+} \frac{a^n}{n!}\mathrm{e}^{-\left(\frac{b-n}{\sqrt{2}\sigma}\right)^2}, \quad n^+ = \lceil n^* + \delta\sigma \rceil, \tag{13}$$

with n^* given by

$$\begin{aligned} n^* &= \sigma\mathcal{W}\left(\frac{a}{\sigma^2}\mathrm{e}^{b/\sigma^2}\right) \\ &\approx \sigma\left(\frac{b}{\sigma^2}\log\left(\frac{a}{\sigma^2}\right) - \log\left(\frac{b}{\sigma^2}\log\left(\frac{a}{\sigma^2}\right)\right)\right) \\ &= \frac{b}{\sigma}\log\left(\frac{a}{\sigma^2}\right) - \sigma\log\left(\frac{b}{\sigma^2}\log\left(\frac{a}{\sigma^2}\right)\right), \end{aligned} \tag{14}$$

where $\mathcal{W}(\cdot)$ denotes the Lambert function. The accuracy of this approximation is controlled by δ. Reference [110] provides a comprehensive analysis on the maximum error value of the truncated sum (13), finding that the infinite sum is well-approximated by taking a manageable number of terms. In particular, we found that for our problem specifications, taking 100 terms was sufficient to ensure an approximation with error less than 0.1 percent.

B. Accelerated Wirtinger Flow With Score-Based Image Prior

For accelerating the WF algorithm, we followed the implementation of [111] as its convergence guarantee was proved. Assuming that the true score function can be learned properly, when we have a trained score function $s_{\boldsymbol{\theta}}(\boldsymbol{x}, \boldsymbol{\sigma})$ by applying (2), the gradient descent algorithm for MAP estimation (1) has the form: $\boldsymbol{x}_{t+1} = \boldsymbol{x}_t - \mu(\nabla g(\boldsymbol{x}_t) + s_{\boldsymbol{\theta}}(\boldsymbol{x}_t, \sigma_k))$. Algorithm 2 summarizes our proposed AWFS algorithm. (Supplement shows the vanilla version without acceleration.) Similar to Langevin dynamics, we choose σ_k to be a descending sequence of noise levels. In practice, we generally use each noise level a fixed

Algorithm 2: Proposed AWFS Method for PR: Accelerated WF with Score-based Image Prior.

Require: Measurement $\boldsymbol{y}$, system matrix $\boldsymbol{A}$, momentum factor $\eta_0 = 1$, step size factor β, truncation operator $\mathcal{P}_C(\cdot) \to [0, C]$; initial image $\boldsymbol{x}_0$, initial auxiliary variables $\boldsymbol{z}_0 = \boldsymbol{w}_0 = \boldsymbol{v}_0 = \boldsymbol{x}_0$, initialize $\sigma_1 > \sigma_2 > \cdots > \sigma_K$.

for $k = 1 : K$ **do**
  for $t = 1 : T$ **do**
    Set step size $\mu = \beta\sigma_k^2$.
    Set $\Delta z_{t,k} = \frac{\eta_{t-1,k}}{\eta_{t,k}}(\boldsymbol{z}_{t,k} - \boldsymbol{x}_{t,k})$.
    Set $\Delta x_{t,k} = \frac{\eta_{t-1,k}-1}{\eta_{t,k}}(\boldsymbol{x}_{t,k} - \boldsymbol{x}_{t-1,k})$.
    Set $\boldsymbol{w}_{t,k} = \mathcal{P}_C(\boldsymbol{x}_{t,k} + \Delta z_{t,k} + \Delta x_{t,k})$.
    Compute $s_{\boldsymbol{\theta}}(\boldsymbol{x}_{t,k}, \sigma_k)$ and $s_{\boldsymbol{\theta}}(\boldsymbol{w}_{t,k}, \sigma_k)$.
    Set $\boldsymbol{z}_{t+1,k} = \boldsymbol{w}_{t,k} - \mu(\nabla g_{\text{PG}}(\boldsymbol{w}_{t,k}) + s_{\boldsymbol{\theta}}(\boldsymbol{w}_{t,k}, \sigma_k))$.
    Set $\boldsymbol{v}_{t+1,k} = \boldsymbol{x}_{t,k} - \mu(\nabla g_{\text{PG}}(\boldsymbol{x}_{t,k}) + s_{\boldsymbol{\theta}}(\boldsymbol{x}_{t,k}, \sigma_k))$.
    Set $\eta_{t+1,k} = \frac{1}{2}(1 + \sqrt{1 + 4\eta_{t,k}^2})$.
    Choose weight factor $\gamma_{t,k}$ (see Theorem 2).
    Set $\boldsymbol{x}_{t+1,k} = \mathcal{P}_C(\gamma_{t,k}\boldsymbol{z}_{t+1,k} + (1 - \gamma_{t,k})\boldsymbol{v}_{t+1,k})$.
  end for
end for
Return $\boldsymbol{x}_{T,K}$.

number of times, with geometrically spaced noise levels between some lower and upper bound. The step size factor β in Algorithm 2 can be selected empirically, but we show that the Lipschitz constant of the gradient $\nabla g_{\text{PG}}(\boldsymbol{x}_t) + s_{\boldsymbol{\theta}}(\boldsymbol{x}_t, \sigma_k)$ exists as demonstrated in Theorem 2 (see proof in Supplement); hence with sufficiently small step size β, the sequence generated by Algorithm 2 will converge to a critical point of the posterior distribution in (1).

We assume that the data allows the neural network to learn the score function well, i.e., $s_{\boldsymbol{\theta}}(\boldsymbol{x}, \sigma) \approx \nabla \log(p_\sigma(\boldsymbol{x}))$, where $p_\sigma(\boldsymbol{x}) = p(\boldsymbol{x}) \circledast \mathcal{N}(0, \sigma^2)$, where $\circledast$ denotes (circular) convolution. Supplement shows that $\nabla \log(p_\sigma(\boldsymbol{x}))$ is Lipschitz continuous on $[-C, C]^N$. Using $p_\sigma(\boldsymbol{x})$, we define the smoothed posterior as

$$p_\sigma(\boldsymbol{x}|\boldsymbol{A}, \boldsymbol{y}, \bar{\boldsymbol{b}}, \boldsymbol{r}) \propto p(\boldsymbol{y}|\boldsymbol{A}, \boldsymbol{x}, \bar{\boldsymbol{b}}, \boldsymbol{r}) p_\sigma(\boldsymbol{x}). \tag{15}$$

Theorem 2: For a smooth density function $p_{\sigma_k}(\boldsymbol{x})$ that has finite expectation with $\sigma_k > 0$, the Lipschitz constant of $\nabla g_{\text{PG}}(\boldsymbol{x}_{t,k}) + s_{\boldsymbol{\theta}}(\boldsymbol{x}_{t,k}, \sigma_k)$ exists when each element in $\boldsymbol{x}_{t,k}$ satisfies $0 \le |x_j| \le C$ for each j. Furthermore, if the weighting factor $\gamma_{t,k}$ is set to 0 if $p_{\sigma_k}(\boldsymbol{v}_{t+1}|\boldsymbol{y}, \boldsymbol{A}, \bar{\boldsymbol{b}}) \le p_{\sigma_k}(\boldsymbol{z}_{t+1}|\boldsymbol{y}, \boldsymbol{A}, \bar{\boldsymbol{b}})$ and 1 otherwise, then with sufficiently small β, the inner iteration sequence $\{\boldsymbol{x}_{t,k}\}_{t=1}^T$ generated by Algorithm 2 is bounded, and any accumulation point of that sequence as $t \to \infty$ is a critical point of the smoothed posterior distribution $p_{\sigma_k}(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{A}, \bar{\boldsymbol{b}}, \boldsymbol{r})$ in (15).

Sketch of Proof: By Lipschitz continuity of $\log(p_\sigma(\boldsymbol{x}))$, and from the design of Algorithm 2, $\boldsymbol{x}_{t,k}$ and $\boldsymbol{w}_{t,k}$ are both bounded between $[-C, C]$ for all t, k, so the Lipschitz constant $\mathcal{L}^*$ of $\nabla g_{\text{PG}}(\cdot) + s_{\boldsymbol{\theta}}(\cdot)$ exists. With the stepsize μ satisfying $0 < \mu < \frac{1}{\mathcal{L}^*}$, and the weighting factor $\gamma_{t,k} \in \{0, 1\}$ being chosen according to the higher posterior probability between $p_{\sigma_k}(\boldsymbol{z}|\boldsymbol{A}, \boldsymbol{y}, \bar{\boldsymbol{b}}, \boldsymbol{r})$ and $p_{\sigma_k}(\boldsymbol{v}|\boldsymbol{A}, \boldsymbol{y}, \bar{\boldsymbol{b}}, \boldsymbol{r})$ (see [111]), we satisfy all conditions in Theorem 1 of [111], which establishes the critical-point convergence of the sequence $\boldsymbol{x}_{t,k}$ generated by Algorithm 2 for any $\sigma_k, k = 1, \ldots, K$. Hence the sequence $\boldsymbol{x}_{t,k}$ generated by Algorithm 2 converges as $t \to \infty$ to a critical-point of the posterior $p_{\sigma_k}(\boldsymbol{x}|\boldsymbol{A}, \boldsymbol{y}, \bar{\boldsymbol{b}}, \boldsymbol{r})$ for any σ_k. Supplement shows the full proof of Theorem 2.

Theorem 2 states that at each iteration, we should choose $\gamma_{t,k} \in \{0, 1\}$ according to the higher posterior probability to ensure convergence. However, we found empirically that setting $\gamma_{t,k} = 0.5$ also led to convergent sequences (see Fig. 9), so we use this practical version for all the results in Section IV. An alternative to our score-based image prior approach would be to alternate data fidelity updates with the sampling step of a diffusion model based on the denoising diffusion probabilistic models (DDPM) framework [75]. Here, we implement and present the results of a slightly altered version (details shown in the Supplement). There are two main changes for this version: first, we initialize with the result of a baseline reconstruction method (in this case, the unregularized Poisson method), rather than pure noise; second, we use the PG likelihood to enforce data fidelity as opposed to the Gaussian likelihood. The former change enables the latter change to be feasible, as it reduces the required number of sampling steps by a factor of 30, allowing for the expensive computation of the PG likelihood each step.

IV. Experiment

A. Experiment Settings

Data: We tested all algorithms on three datasets: 162 histopathology images related to breast cancer [112] (train/val/test is 122/20/20); 920 images from CelebA dataset [113] (train/val/test is 800/100/20); and 720 images from a homemade CT-density dataset (train/val/test is 600/100/20). The CT-density dataset was generated from single-photon emission computerized tomography (SPECT)/CT images for Yttrium-90 radionuclide therapy after applying the CT-to-density calibration curve [114]. Although the size of training datasets are relatively small compared to typical datasets such as ImageNet or LSUN [65], [67] that have millions of images, we do not require the score functions to learn image priors strong enough to generate realistic images from white Gaussian noise; rather, it suffices for the priors to be able to denoise moderately noisy images.

System Model. The system matrix is based on the discrete Fourier transform of the concatenation of the true image $\boldsymbol{x}$, a blank image $\boldsymbol{0}$ and a reference image $\boldsymbol{r}$ with scaling and oversampling. We set the scaling factor α to be in the range $[0.02, 0.035]$ so that the average counts per pixel range from 6 to 25; the oversampled ratio is set to 2. We set $\boldsymbol{r}$ to be a binary random image similar to what was used in [39]. The standard deviation of the Gaussian read noise added to the measurements $\boldsymbol{y}$ was set as $\sigma \in [0.5, 1.5]$.

Algorithms: For unregularized algorithms, we implemented Gaussian WF [33], Poisson WF [81] and Poisson-Gaussian

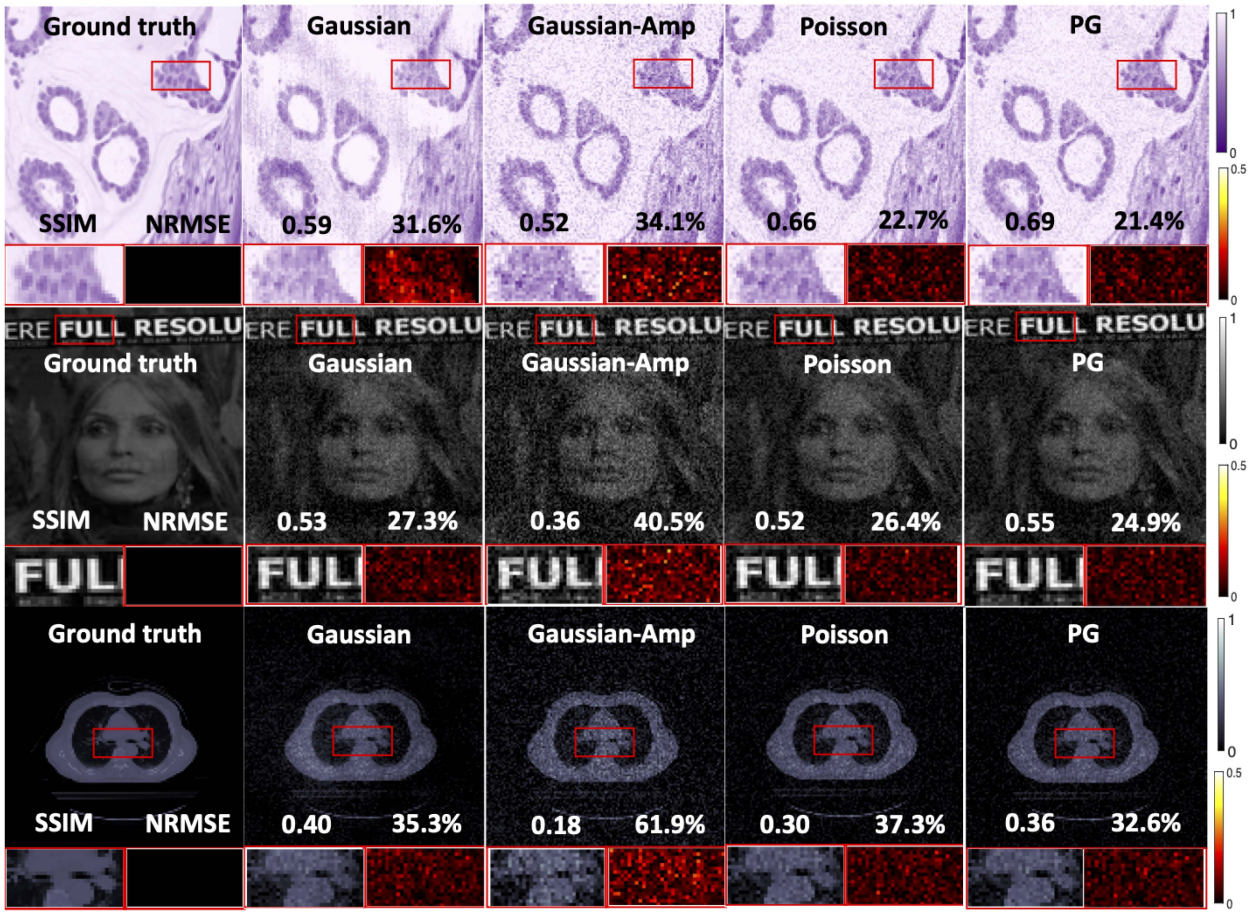


Fig. 2. Reconstructed images by unregularized methods (Gaussian, Gaussian-Amplitude, Poisson and Poisson-Gaussian) on Histopathology dataset [112], celebA dataset [113] and CT-density dataset. The bottom left/right subfigures correspond to the zoomed in area and the error map for each image. We used $\alpha = 0.035$ and $\sigma = 1$.

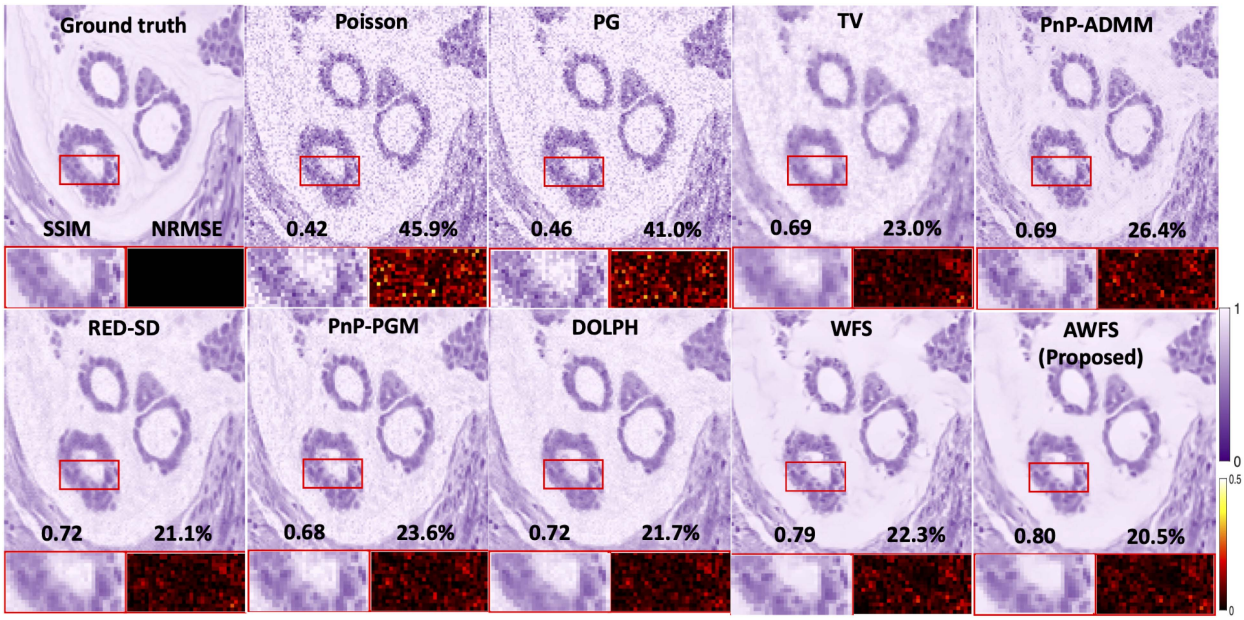


Fig. 3. Reconstructed images on dataset [112]. The bottom left/right subfigures correspond to the zoomed in area and the error map for each image. We used $\alpha = 0.02$ and $\sigma = 1$.

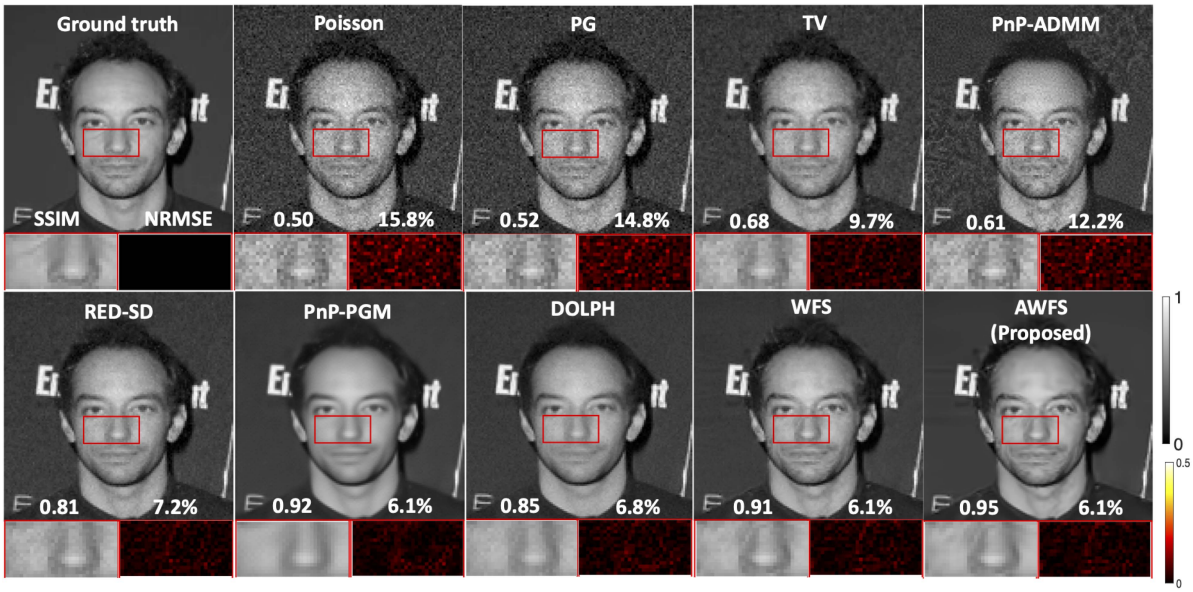


Fig. 4. Reconstructed images on celebA dataset [113]. The bottom left/right subfigures correspond to the zoomed in area and the error map for each image. We used $\alpha = 0.035$ and $\sigma = 1$.

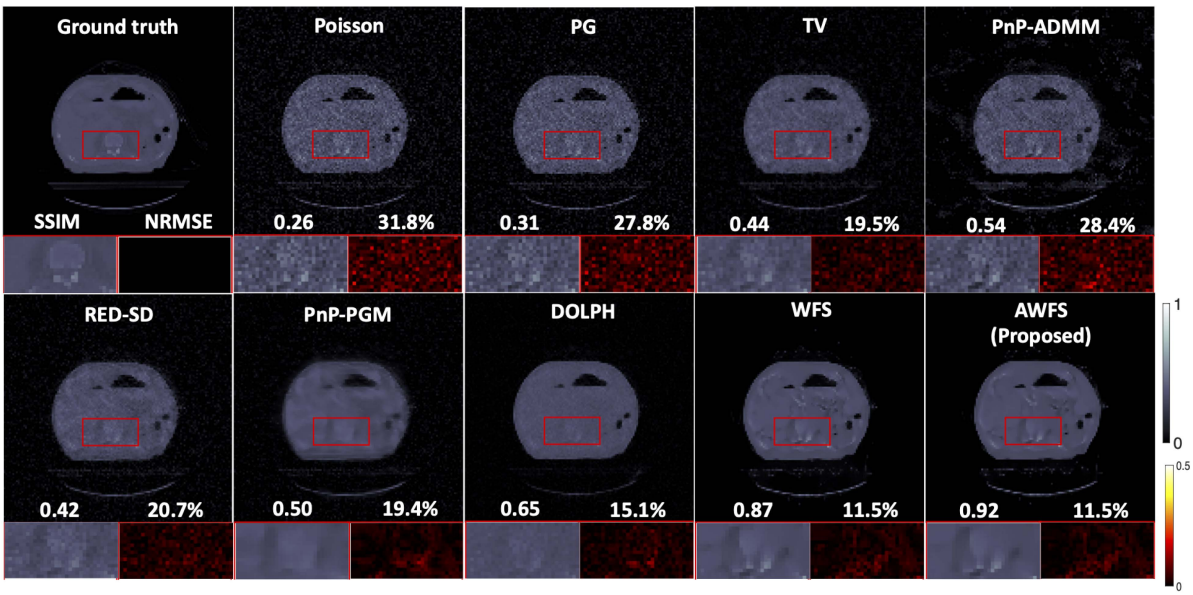


Fig. 5. Reconstructed images on CT-density dataset. The bottom left/right subfigures correspond to the zoomed in area and the error map for each image. We used $\alpha = 0.035$ and $\sigma = 1$.

WF. For regularized algorithms, we implemented smoothed total variation (TV) based on the Huber function [115, p. 184] and PnP/RED methods with the DnCNN denoiser [116]: PnP-ADMM [101], PnP-PGM [102], and RED-SD [58]. We also implemented the RED-SD algorithm with "Noise2Self" zero-shot image denoising network [60] (RED-SD-SELF). For diffusion models, we implemented DOLPH [75] and our proposed AWFS. Supplement shows the implementation details of each algorithm. We used spectral initialization [77] for the Gaussian PR and Poisson PR methods; we then used the output results from Poisson PR to initialize other algorithms. We ran all algorithms until the normalized root mean squared error (NRMSE) between consecutive iterations differed by less than 0.01 percent or reached the maximum number of iterations (e.g., 50).

To evaluate the robustness and limitation of these algorithms, we first tuned the parameters for each algorithm at the noise level when $\alpha = 0.030$ and $\sigma = 1$, and then held them fixed throughout all experiments (Tables I, II, Figs. 7 and 8). In practice the ground truths are unknown, so oracle tuning of test datasets is infeasible (though some form of cross validation may be possible). Though the numbers reported could fluctuate after careful refinement, e.g., by performing grid search on tuning parameters, such techniques would potentially impede the algorithm's practical use.

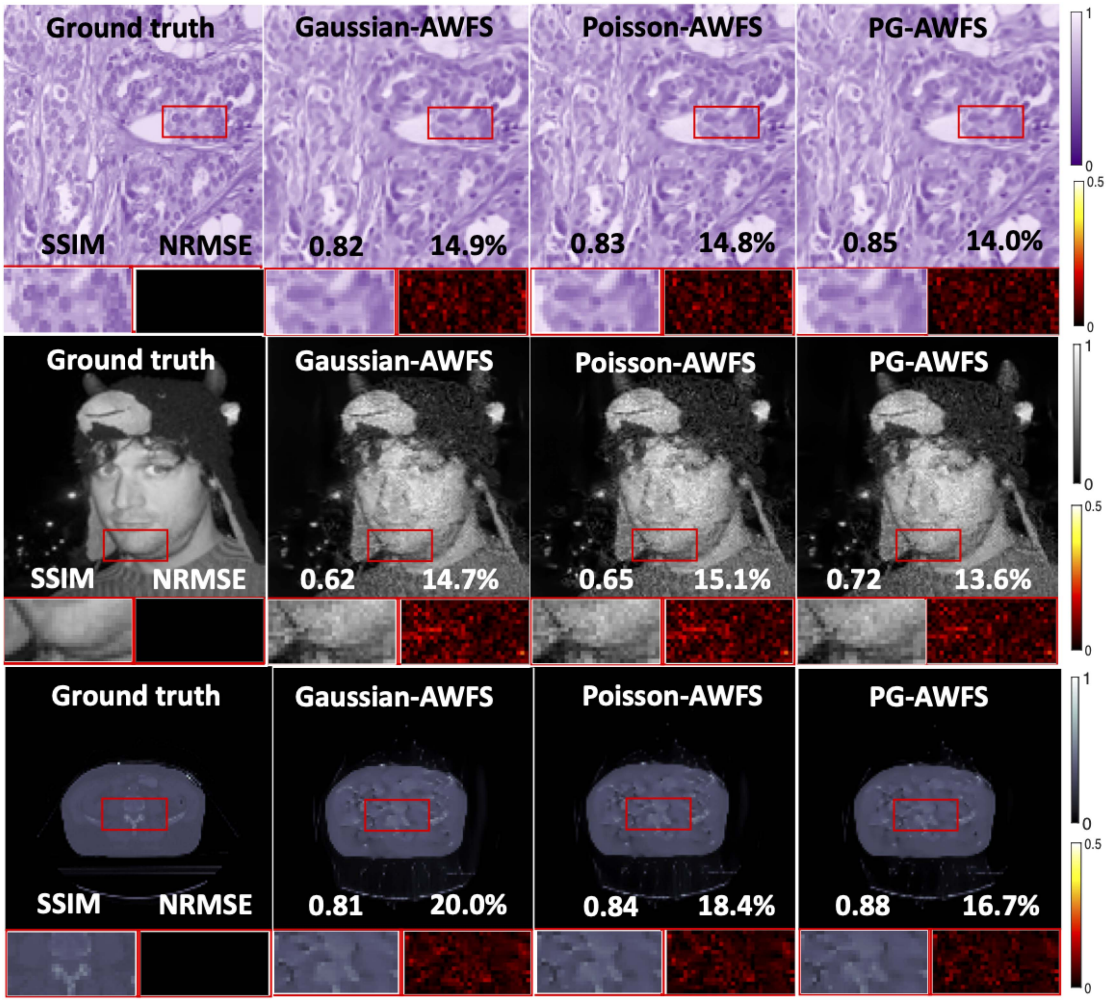


Fig. 6. Reconstructed images by Gaussian, Poisson and Poisson-Gaussian log-likelihood model with AWFS image prior. Tested on Histopathology dataset [112], celebA dataset [113] and CT-density dataset. The bottom left/right subfigures correspond to the zoomed in area and the error map for each image. α and σ were set to 0.025 and 1, respectively.

TABLE I
SSIM AND NRMSE FOR POISSON AND POISSON-GAUSSIAN LIKELIHOODS

Likelihood	Unregularized (SSIM/NRMSE)		DOLPH (SSIM/NRMSE)		AWFS (SSIM/NRMSE)	
DataSet: Histopathology [112]						
Gaussian [33]	0.52 ± 0.18	41.2 ± 25.3	0.76 ± 0.07	18.0 ± 3.0	0.84 ± 0.06	16.2 ± 3.7
Poisson [81]	0.54 ± 0.18	31.7 ± 10.2	0.72 ± 0.13	19.5 ± 6.1	0.83 ± 0.06	16.2 ± 3.7
Poisson-Gaussian	0.57 ± 0.18	28.9 ± 9.0	0.80 ± 0.06	16.0 ± 2.9	**0.85 ± 0.05**	**15.4 ± 3.7**
DataSet: CelebA [113]						
Gaussian [33]	0.31 ± 0.09	55.6 ± 13.9	0.70 ± 0.12	14.5 ± 17.4	0.72 ± 0.16	15.3 ± 11.8
Poisson [81]	0.39 ± 0.10	24.5 ± 11.4	0.61 ± 0.12	15.6 ± 10.6	0.72 ± 0.16	15.2 ± 11.8
Poisson-Gaussian	0.42 ± 0.10	21.8 ± 9.1	0.71 ± 0.11	**13.7 ± 11.1**	**0.74 ± 0.15**	14.8 ± 11.9
DataSet: CT-Density						
Gaussian [33]	0.29 ± 0.09	50.5 ± 8.0	0.51 ± 0.12	22.4 ± 3.9	0.82 ± 0.11	19.1 ± 4.8
Poisson [81]	0.19 ± 0.06	48.9 ± 13.1	0.38 ± 0.11	25.6 ± 7.5	0.84 ± 0.08	17.8 ± 4.3
Poisson-Gaussian	0.24 ± 0.06	40.8 ± 9.5	0.55 ± 0.08	20.0 ± 3.3	**0.88 ± 0.05**	**16.4 ± 3.7**

Results were averaged across 7 different noise levels by varying $\alpha \in 0.02 : 0.005 : 0.035$ in (3), with $\sigma = 1$.

TABLE II
SSIM AND NRMSE USING POISSON GAUSSIAN LIKELIHOOD WITH DIFFERENT REGULARIZATION/IMAGE PRIOR APPROACHES

Dataset	Histopathology [112]		CelebA [113]		CT-Density	
Methods	SSIM	NRMSE (%)	SSIM	NRMSE (%)	SSIM	NRMSE (%)
Unregularized	0.57 ± 0.18	28.9 ± 9.0	0.42 ± 0.10	21.8 ± 9.1	0.24 ± 0.06	40.8 ± 9.5
RED-SD-SELF [60]	0.66 ± 0.13	21.9 ± 4.5	0.60 ± 0.09	15.9 ± 10.6	0.34 ± 0.04	28.1 ± 4.1
PnP-ADMM [101]	0.71 ± 0.11	20.7 ± 4.2	0.56 ± 0.08	16.7 ± 8.1	0.55 ± 0.03	31.2 ± 2.7
TV regularizer	0.72 ± 0.11	18.2 ± 3.9	0.64 ± 0.07	14.4 ± 8.6	0.41 ± 0.03	23.7 ± 2.8
RED-SD [58]	0.76 ± 0.09	16.8 ± 3.6	0.69 ± 0.11	13.9 ± 10.9	0.38 ± 0.04	25.9 ± 4.0
PnP-PGM [102]	0.78 ± 0.11	16.5 ± 4.5	**0.74 ± 0.14**	**13.5 ± 11.3**	0.42 ± 0.07	24.6 ± 4.4
DOLPH [75]	0.80 ± 0.06	16.0 ± 2.9	0.71 ± 0.11	13.7 ± 11.1	0.55 ± 0.08	20.0 ± 3.3
WFS*	0.76 ± 0.12	18.2 ± 5.5	0.63 ± 0.16	16.9 ± 11.8	0.53 ± 0.17	21.3 ± 7.6
WFS†	0.83 ± 0.06	16.2 ± 4.0	0.70 ± 0.16	15.7 ± 11.8	0.74 ± 0.13	17.3 ± 4.8
AWFS (Proposed)	**0.85 ± 0.05**	**15.4 ± 3.7**	**0.74 ± 0.15**	14.8 ± 11.9	**0.88 ± 0.05**	**16.4 ± 3.7**

Resultswere averaged across 7 different noise levels by varying $\alpha \in 0.02 : 0.005 : 0.035$ in (3). WFS* runs the same number of iterations as AWFS whereas WFS† runs more iterations until convergence.

Network Training: For PnP denoising networks, we trained all denoisers on different noise levels $\sigma \in \{9, 11, 13, 15\}$ and found that $\sigma = 15$ worked the best on our data. We also used the denoiser scaling technique from [117] to dynamically adjust the performance of all PnP methods. To perform score matching, we applied 20 geometrically spaced noise levels between 0.005 and 0.1 on each of the training images. All networks were implemented in PyTorch and trained on an NVIDIA Quadro RTX 5000 GPU using the ADAM optimizer [118] for 1000 epochs with the best one being selected based on the mean squared error (MSE) validation loss.

B. Results

We compared all implemented algorithms both qualitatively, by visualizing the reconstructed images and residual errors, and quantitatively, by computing the NRMSE and structural similarity index measure (SSIM) [119]. Due to the global phase ambiguity, i.e., all the algorithms can recover the signal only to within a constant phase shift due to the loss of global phase information, we corrected the phase of $\hat{\boldsymbol{x}}$ by $\hat{\boldsymbol{x}}_{\text{corrected}} \triangleq \text{sign}(\langle \hat{\boldsymbol{x}}, \boldsymbol{x}_{\text{true}} \rangle)\hat{\boldsymbol{x}}$.

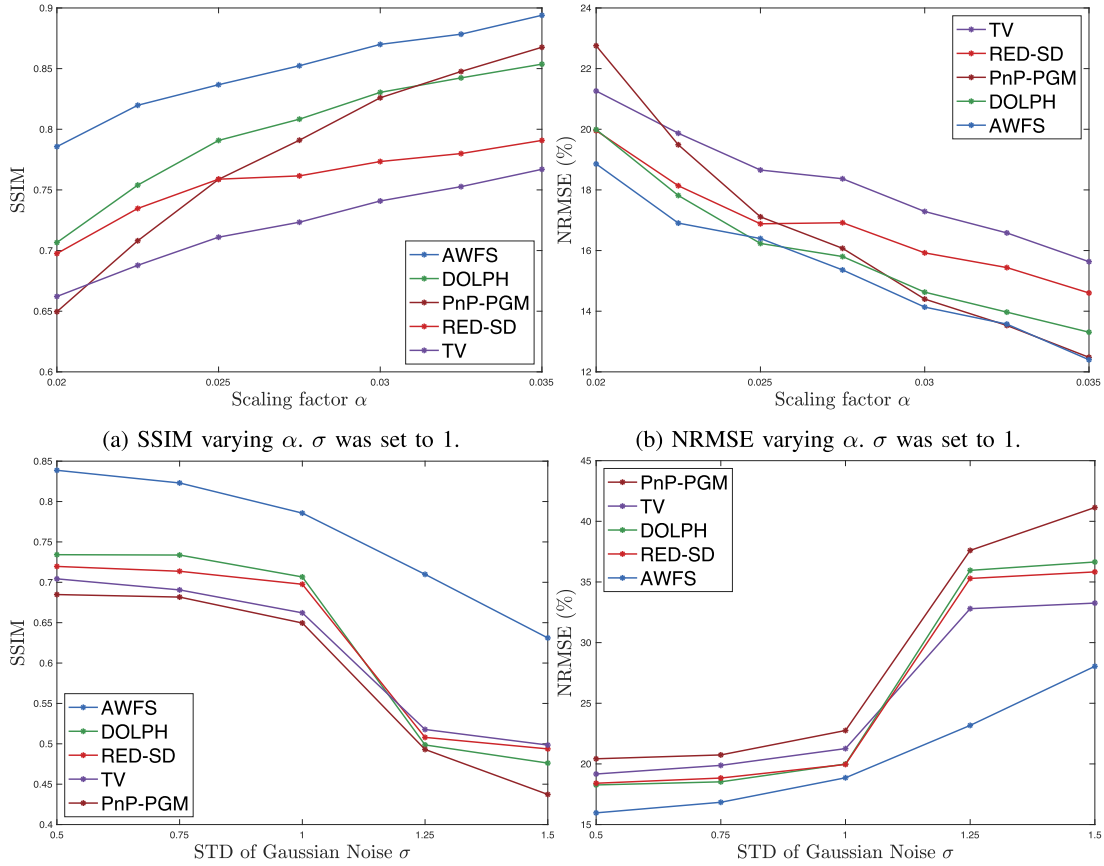
Fig. 2 shows experiments of running unregularized methods based on different noise models on the histopathology, CelebA, and CT density datasets. For comparison, we ran the unregularized methods with a Gaussian only noise model, Poisson only, and PG noise model.

Figs. 3, 4, and 5 visualize reconstructed images generated by algorithms mentioned in the previous section. The WF with PG likelihood outperforms WF with Poisson likelihood with a consistently higher SSIM and lower NRMSE. Of the regularized algorithms with PG likelihood, our proposed AWFS had less visual noise and achieved greater detail recovery compared to other methods, as evidenced by the zoomed-in area in these figures. Fig. 6 shows that for a variety of datasets, when combined with the AWFS method, while the Poisson only and Gaussian only models lead to reasonable reconstructions, the PG noise model leads to the highest quality image. For all three datasets shown, when used in conjunction with our AWFS method, including both Poisson and Gaussian likelihoods results in the highest quality reconstruction both in terms of quantitative metrics as well as visually. Thus, although the score function provides a useful prior for recovering an image when the measurement is very noisy, a proper noise model is also crucial to a high quality reconstruction.

For quantitative evaluations, Table I illustrates the effect of using our proposed PG likelihood as compared to the simpler Poisson likelihoods. We did not run the Gaussian likelihood with DOLPH or AWFS due to the abysmal performance with this likelihood. In all cases, usage of the PG likelihood results in improved image quality in terms of both metrics. Table II consists of experiments using the PG likelihood and shows the efficacy of the proposed AWFS method over other methods. In particular, our AWFS had superior quantitative performance over all other compared methods on the histopathology and CT-density datasets; in contrast, the PnP-Prox showed the lowest NRMSE on celebA dataset. This is likely due to higher randomness in celebrity faces because the effectiveness of generative models can vary depending on the dataset used. Thus, when provided with a small amount of training data with high randomness, image denoising models (DnCNN) may be more effective than generative models.

We also tested the robustness of the leading algorithms in Table II, by varying both scaling factor α and STD of Gaussian noise σ. Figs. 7 and 8 illustate results, where our AWFS algorithm had the highest SSIM and lowest NRMSE. In Fig. 8, AWFS demonstrated minimal variations in SSIM and NRMSE metrics than DOLPH as evidenced by the smaller discrepancies in SSIM (0.17vs. 0.23) and NRMSE (12.6% vs. 18.2%) when σ varies from 0.75 to 1.5. Fig. 9 compares the convergence rate of AWF vs. WF for the Poisson and PG likelihood, respectively. Under a variety of noise levels, AWF consistently converged faster than WF in terms of number of iterations.

It is a known property of diffusion models that they can produce images with hallucinated features if the measurements are insufficiently informative. In the case of low-count phase retrieval with serious corruptions of both Poisson and Gaussian noise, as is investigated here, the measurement is highly corrupted and contains magnitude-only measurements of the



(a) SSIM varying α. σ was set to 1.

(b) NRMSE varying α. σ was set to 1.

(c) SSIM varying σ; α was set as 0.02.

(d) NRMSE varying σ; α was set as 0.02.

Fig. 7. Comparison of SSIM and NRMSE varying scaling factor $\alpha \in [0.02, 0.035]$ and STD of Gaussian noise $\sigma \in [0.25, 1.5]$ defined in (3).

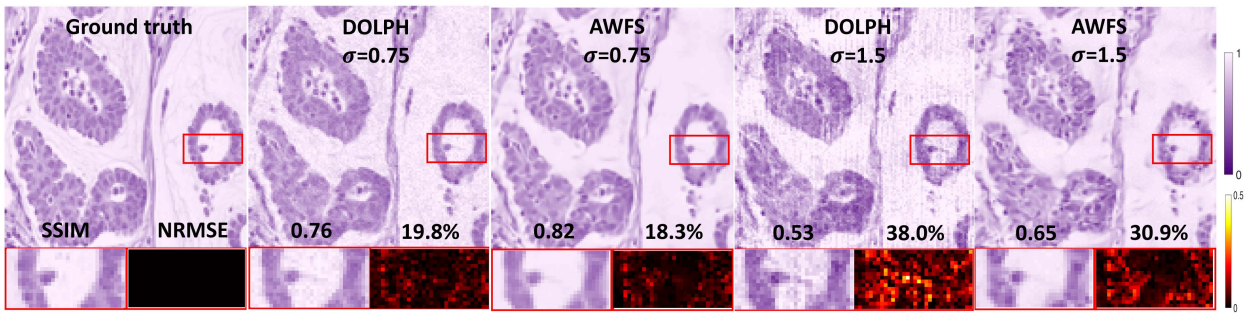
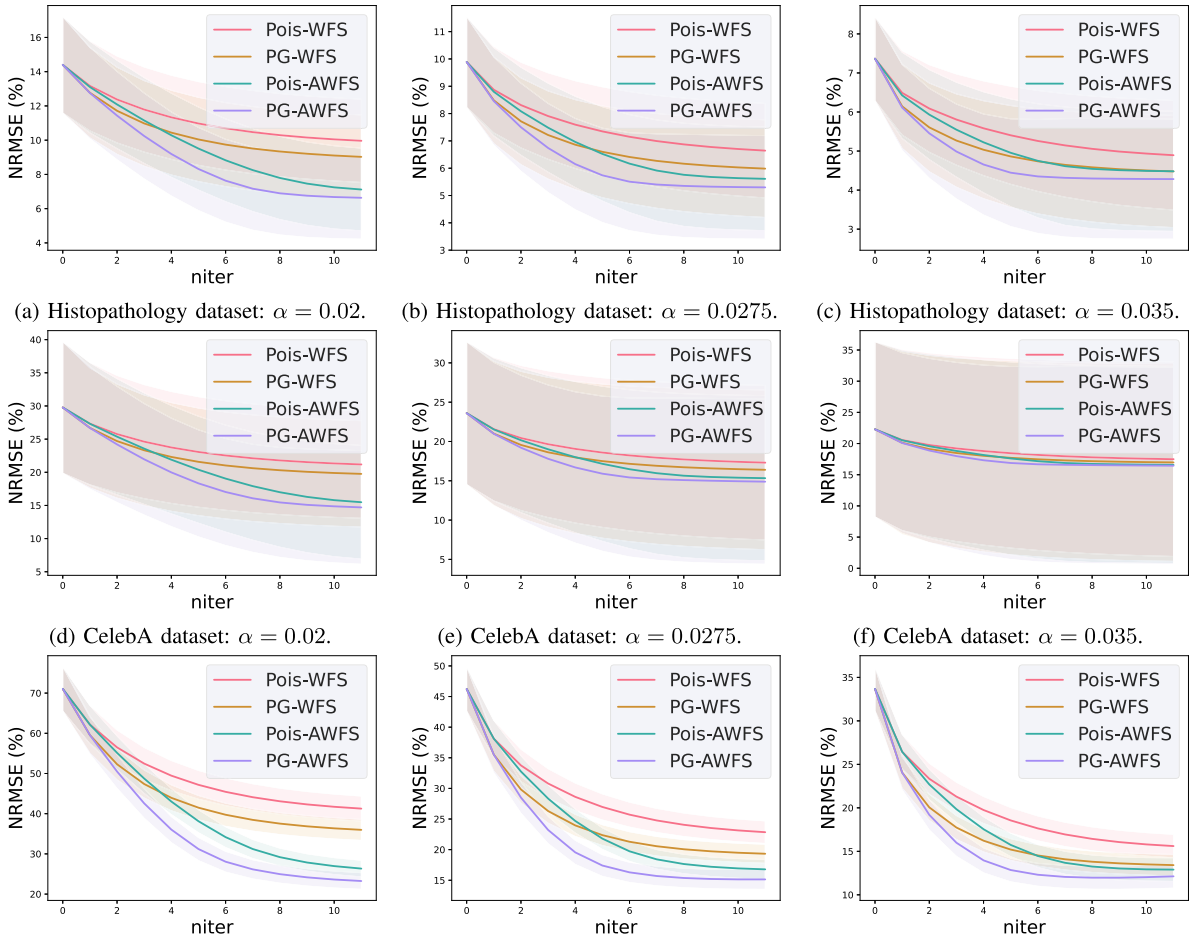


Fig. 8. Reconstructed images by DOLPH [75] and our proposed AWFS method under different σ values, for $\alpha = 0.02$.

original signal. Thus, it may be difficult for the diffusion models to avoid some otherwise realistic hallucinations if the data consistency is not strong enough to guide the model away from such hallucinations. On the other hand, if the measurements are less corrupted, then the data consistency should be strong enough to avoid such hallucinations. Fig. 10 provides examples of this for the CT image dataset via a comparison of the reconstruction quality of the AWFS method over a range of count levels. With the lowest scaling factor, e.g., $\alpha = 0.02$, the measurements were seriously corrupted with noise, and the method may hallucinate some features. However, at higher count level, e.g., $\alpha = 0.05$, there is enough information in the measurement to enforce consistency and avoid noticeable hallucinations. We performed the same experiment twice with different noisy initializations and all other parameters held equal to demonstrate robustness of the method under different initializations.

V. Discussion

PR has a long-standing history in the field of signal processing and imaging. Pioneering works such as the error reduction (ER) and hybrid input-output (HIO) algorithms by Gerchberg Saxton [25] and Fienup [90] have been proposed to address this problem. These iterative algorithms involve constraints imposed on evaluations between the image domain and frequency domain. However, these methods have limitations in terms of the



(a) Histopathology dataset: $\alpha = 0.02$. (b) Histopathology dataset: $\alpha = 0.0275$. (c) Histopathology dataset: $\alpha = 0.035$.

(d) CelebA dataset: $\alpha = 0.02$. (e) CelebA dataset: $\alpha = 0.0275$. (f) CelebA dataset: $\alpha = 0.035$.

(g) CT-density dataset: $\alpha = 0.02$. (h) CT-density dataset: $\alpha = 0.0275$. (i) CT-density dataset: $\alpha = 0.035$.

Fig. 9. Comparing AWF vs. WF with NRMSE vs. number of iterations under different noise levels. The curves and shadows represent the mean and standard deviation, respectively.

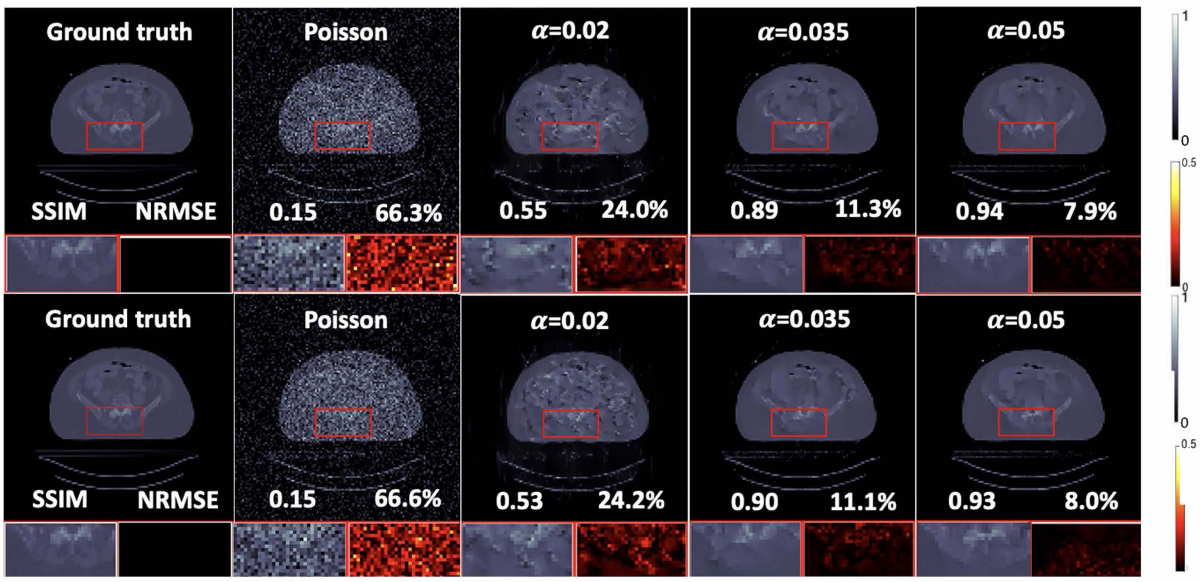


Fig. 10. Reconstructed images by the unregularized Poisson method (the second column) as well as with the AWFS method for different scaling factors α (third to fifth columns). The top and bottom rows show reconstructions from different measurement realizations.

quality of reconstructed images and their convergence remains uncertain [120]. Another approach to solving PR problems is through compressed sensing and optimization techniques like Wirtinger flow (WF) [33], matrix lifting [5], [19], [20], MM [24] and ADMM [30]. This paper focuses on the WF algorithm due to being straightforward to incorporate with the DL regularizer for the image prior. The likelihood modelling of the noise statistics existing in the measurement is also critical. Previous studies have primarily focused on modelling either Gaussian or Poisson likelihood only, but in practical scenarios, both types of noise are often encountered. Therefore, this paper contributes to a more practical perspective of addressing the holographic PR problem by using a PG likelihood and incorporating state-of-the-art deep learning image priors. In the case where the measurement is contaminated with Poisson and Gaussian noise, the speedup in reconstruction is crucial, as the bottleneck of our algorithm is in computing the PG likelihood. Additionally, though it is viable to perform a large number of neural network evaluations to perform image reconstruction, it is unrealistic to compute a similarly large number of PG likelihoods. Thus, we perform acceleration in WF algorithm following [111], which guarantees convergence to a critical point for the Holographic PR problems.

In our evaluation of three datasets, we consistently observed that the use of PG likelihood yielded superior performance compared to using either Poisson or Gaussian likelihood alone, as expected. Additionally, the results obtained from the CT-density dataset were generally of lower quality than those from the other two datasets. This can be attributed to lower average counts per pixel (many zero pixels near the image borders). Using a DL image prior can be considered from two perspectives: training a denoiser or training to learn the density distribution of images. In our work, we applied both approaches and observed that the effectiveness of these methods differed depending on the dataset tested. Specifically, in the Histopathology dataset [112] and the CT-density dataset, where the images share similar structures, the generative models performs better even when trained with limited data. In the case of the CelebA dataset [113], which includes a wide variety of celebrity faces, generative models did not exhibit as strong performance as denoiser methods when trained on limited data. This is likely due to the fact that generating high-quality images is generally more challenging than removing noise from existing images and may necessitate a larger training dataset. The effectiveness of accelerated WF compared to vanilla WF is due to the non-convexity of the PR problem. Although recent advances in geometric landscape analysis of PR can guarantee that all local minimizers are global even with random initialization [121], in practice the measurements are contaminated by noise so that many more measurements are required for the cost function to have a benign geometric landscape.

Despite the promising results achieved with our proposed AWFS approach, there are several limitations of our work. First, the approximate calculation of the infinite sum in (4) is accurate but computationally expensive. Future work should seek ways to accelerate this calculation while maintaining accuracy. Second, we did not implement and test the accelerated WF applied on the diffusion posterior sampling (DPS) method [109], for which the network is fine-tuned from a pretrained state-of-the-art diffusion model. This approach has the potential to advance current methods in PR problem and we will investigate it in the future. Another limitation of the proposed method is that it has been demonstrated on measurements that are based on simulations. To further demonstrate the efficacy of the method in a real-world setting, future work should consist of evaluating the accuracy of the methods when run on real measurement data. Finally, our experiments are limited to real-valued images, however, our method can be extended to handle complex-values images by splitting real and imaginary components into separate reconstruction routines with different pretrained neural networks [78]. Addressing these limitations will be the future direction of this work.

VI. Conclusion

We proposed a novel algorithm based on **A**ccelerated **W**irtinger **F**low and **S**core-based image prior (AWFS) for Poisson-Gaussian holographic phase retrieval. Simulation experiments demonstrate that the proposed AWFS method produced the best reconstruction quality both qualitatively and quantitatively and was more robust to various noise levels, compared to other state-of-the-art methods. Furthermore, we proved that our proposed algorithm has a critical-point convergence guarantee. Therefore, after the method is extended to accommodate complex-valued images, it should have much promise for real-world PR applications.

References

[1] J. Hu, Z. Li, X. Xu, L. Shen, and J. A. Fessler, "Poisson-Gaussian holographic phase retrieval with score-based image prior," in *Proc. NeurIPS Workshop Deep Learn. Inverse Problems*, 2023. [Online]. Available: https://openreview.net/forum?id=j6YOCS0cyM

[2] K. Jaganathan, Y. C. Eldar, and B. Hassibi, "Phase retrieval: An overview of recent developments," in *Optical Compressive Imaging*, 1st edn. ImprintCRC Press, 2016.

[3] J. C. Dainty and J. R. Fienup, *Image Recovery: Theory and Application*, H. Stark, ed., Academic Press, 1987, 231–275, Ch. 7.

[4] R. P. Millane, "Phase retrieval in crystallography and optics," *J. Opt. Soc. Amer. A*, vol. 7, no. 3, pp. 394–411, Mar. 1990.

[5] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev, "Phase retrieval with application to optical imaging: A contemporary overview," *IEEE Signal Process. Mag.*, vol. 32, no. 3, pp. 87–109, May 2015.

[6] L. Bian et al., "Fourier ptychographic reconstruction using poisson maximum likelihood and truncated Wirtinger gradient," *Nat. Sci. Rep.*, vol. 6, no. 1, 2016, Art. no. 27384 .

[7] Y. Zhang, P. Song, and Q. Dai, "Fourier ptychographic microscopy using a generalized anscombe transform approximation of the mixed Poisson-Gaussian likelihood," *Opt. Exp.*, vol. 25, no. 1, pp. 168–79, Jan. 2017.

[8] R. Xu et al., "Accelerated wirtinger flow: A fast algorithm for ptychography," 2018, *arXiv:1806.05546*.

[9] X. Tian, "Fourier ptychographic reconstruction using mixed gaussian-poisson likelihood with total variation regularisation," *Electron. Lett.*, vol. 55, no. 19, pp. 1041–1043, Sep. 2019.

[10] T. Latychevskaia, "Iterative phase retrieval in coherent diffractive imaging: Practical issues," *Appl. Opt.*, vol. 57, no. 25, pp. 7187–7197, 2018.

[11] D. A. Barmherzig, J. Sun, E. J. Candes, T. Lane, and P.-N. Li, "Dual-reference design for holographic phase retrieval," in *2019 13th Int. Conf. Sampling Theory Appl.*, 2019, pp. 1–4.

[12] D. A. Barmherzig, J. Sun, P.-N. Li, T. Lane, and E. J. Candés, "Holographic phase retrieval and reference design," *Inverse Problems*, vol. 35, no. 9, 2019, Art. no. 094001.

[13] M. Saliba, T. Latychevskaia, J. Longchamp, and H. Fink, "Fourier transform holography: A lensless non-destructive imaging technique," *Microsc. Microanalysis*, vol. 18, no. S2, pp. 564–565, 2012.
[14] M. Guizar-Sicairos and J. R. Fienup, "Holography with extended reference by autocorrelation linear differential operation," *Opt. Exp.*, vol. 15, no. 26, pp. 17592–17612, 2007. [Online]. Available: https://opg.optica.org/oe/abstract.cfm?URI=oe-15-26-17592
[15] L. Wang, "The poisson channel with varying dark current known to the transmitter," *IEEE Trans. Inf. Theory*, vol. 65, no. 8, pp. 4966–4978, Aug. 2019.
[16] V. Mannam et al., "Real-time image denoising of mixed Poisson–Gaussian noise in fluorescence microscopy images using imageJ," *Optica*, vol. 9, no. 4, pp. 335–345, 2022. [Online]. Available: https://opg.optica.org/optica/abstract.cfm?URI=optica-9-4-335
[17] B. Jiang, K. Meng, and K. Youcef-Toumi, "Quantification and reduction of Poisson-Gaussian mixed noise induced errors in ellipsometry," *Opt. Exp.*, vol. 29, no. 17, pp. 27057–27070, 2021. [Online]. Available: https://opg.optica.org/oe/abstract.cfm?URI=oe-29-17-27057
[18] S. Yang and B. Lee, "Poisson-gaussian noise reduction using the hidden Markov model in contourlet domain for fluorescence microscopy images," *PLoS One*, vol. 10, no. 9, 2015, Art. no. e0136964.
[19] E. J. Candés, T. Strohmer, and V. Voroninski, "PhaseLift: Exact and stable signal recovery from magnitude measurements via convex programming," *Commun. Pure Appl. Math.*, vol. 66, no. 8, pp. 1241–1274, 2013.
[20] E. J. Candes, Y. C. Eldar, T. Strohmer, and V. Voroninski, "Phase retrieval via matrix completion," *SIAM J. Imag. Sci.*, vol. 6, no. 1, pp. 199–225, 2013.
[21] X. Jiang, S. Rajan, and X. Liu, "Wirtinger flow method with optimal stepsize for phase retrieval," *IEEE Signal Process. Lett.*, vol. 23, no. 11, pp. 1627–1631, Nov. 2016.
[22] T. T. Cai, X. Li, and Z. Ma, "Optimal rates of convergence for noisy sparse phase retrieval via thresholded wirtinger flow," *Ann. Stat.*, vol. 44, no. 5, pp. 2221–2251, 2016.
[23] M. Soltanolkotabi, "Structured signal recovery from quadratic measurements: Breaking sample complexity barriers via nonconvex optimization," *IEEE Trans. Inf. Theory*, vol. 65, no. 4, pp. 2374–2400, Apr. 2019.
[24] T. Qiu, P. Babu, and D. P. Palomar, "PRIME: Phase retrieval via majorization-minimization," *IEEE Trans. Sigal Process.*, vol. 64, no. 19, pp. 5174–86, Oct. 2016.
[25] R. W. Gerchberg and W. O. Saxton, "Practical algorithm for determination of phase from image and diffraction plane pictures," *OPTIK*, vol. 35, no. 2, pp. 237–246, 1972.
[26] P. Netrapalli, P. Jain, and S. Sanghavi, "Phase retrieval using alternating minimization," *IEEE Trans. Signal Process.*, vol. 63, no. 18, pp. 4814–4826, Sep. 2015.
[27] I. Waldspurger, "Phase retrieval with random gaussian sensing vectors by alternating projections," *IEEE Trans. Inf. Theory*, vol. 64, no. 5, pp. 3301–3312, May 2018.
[28] B. Gao and Z. Xu, "Phaseless recovery using the gauss-newton method," *IEEE Trans. Signal Process.*, vol. 65, no. 22, pp. 5885–5896, Nov. 2017.
[29] L. Ji and Z. Tie, "On gradient descent algorithm for generalized phase retrieval problem," in *2016 IEEE 13th Int. Conf. Signal Process.*, 2016, pp. 320–325.
[30] J. Liang, P. Stoica, Y. Jing, and J. Li, "Phase retrieval via the alternating direction method of multipliers," *IEEE Signal Process. Lett.*, vol. 25, no. 1, pp. 5–9, Jan. 2018.
[31] Y. Yang, M. Pesavento, Y. C. Eldar, and B. Ottersten, "Parallel coordinate descent algorithms for sparse phase retrieval," in *2019 IEEE Int. Conf. Acoust., Speech Signal Process.*, 2019, pp. 7670–7674.
[32] J. Sun, Q. Qu, and J. Wright, "A geometric analysis of phase retrieval," in *2016 IEEE Int. Symp. Inf. Theory*, 2016, pp. 2379–2383.
[33] E. Candes, X. Li, and M. Soltanolkotabi, "Phase retrieval via wirtinger flow: Theory and algorithms," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 1985–2007, Apr. 2015.
[34] Z. Wu, Y. Sun, J. Liu, and U. Kamilov, "Online regularization by denoising with applications to phase retrieval," in *2019 IEEE/CVF Int. Conf. Comput. Vis. Workshop*, 2019, pp. 3887–3895.
[35] P. Thibault and M. Guizar-Sicairos, "Maximum-likelihood refinement for coherent diffractive imaging," *New J. Phys.*, vol. 14, no. 6, Jun. 2012, Art. no. 063004.
[36] A. Goy, K. Arthur, S. Li, and G. Barbastathis, "Low photon count phase retrieval using deep learning," *Phys. Rev. Lett.*, vol. 121, no. 24, Dec. 2018, Art. no. 243902.
[37] D. A. Barmherzig and J. Sun, "Low-photon holographie phase retrieval," in *Imag. Appl. Opt. Congr.*, 2020.
[38] I. Vazquez, I. E. Harmon, J. C. R. Luna, and M. Das, "Quantitative phase retrieval with low photon counts using an energy resolving quantum detector," *J. Opt. Soc. Amer. A*, vol. 38, no. 1, pp. 71–79, Jan. 2021.
[39] H. Lawrence, D. Barmherzig, H. Li, M. Eickenberg, and M. Gabrie, "Phase retrieval with holography and untrained priors: Tackling the challenges of low-photon nanoscale imaging," in *Proc. 2nd Math. Sci. Mach. Learn. Conf.*, 2022, vol. 145, pp. 516–567.
[40] Z. Li and J. A. Fessler, "Uniform Cramér–Rao lower bound for phase retrieval: An empirical study," in *2022 56th Asilomar Conf. Signals, Syst., Comput.*, 2022, pp. 111–115.
[41] A. Gnanasambandam and S. H. Chan, "Image classification in the dark using quanta image sensors," in *Computer Vis. – ECCV 2020, 16th Eur. Conf.*, 2020, pp. 484–501, doi: 10.1007/978-3-030-58598-3_29.
[42] Z. Li, K. Lange, and J. A. Fessler, "Poisson phase retrieval in very low-count regimes," *IEEE Trans. Comput. Imag.*, vol. 8, pp. 838–850, 2022.
[43] G. Fatima, Z. Li, A. Arora, and P. Babu, "PDMM: A novel primal-dual majorization-minimization algorithm for poisson phase-retrieval problem," *IEEE Trans. Signal Process.*, vol. 70, pp. 1241–1255, 2022.
[44] Y. Chen and E. J. Candes, "Solving random quadratic systems of equations is nearly as easy as solving linear systems," *Commun. Pure Appl. Math.*, vol. 70, no. 5, pp. 822–83, May 2017.
[45] H. Zhang, Y. Liang, and Y. Chi, "A nonconvex approach for phase retrieval: Reshaped wirtinger flow and incremental algorithms," *J. Mach. Learn. Res.*, vol. 18, no. 141, pp. 1–35, 2017. [Online]. Available: http://jmlr.org/papers/v18/16-572.html
[46] H. Chang, Y. Lou, Y. Duan, and S. Marchesini, "Total variation–based phase retrieval for poisson noise removal," *SIAM J. Imag. Sci.*, vol. 11, no. 1, pp. 24–55, 2018.
[47] D. L. Snyder, A. M. Hammoud, and R. L. White, "Image recovery from data acquired with a charge-coupled-device camera," *J. Opt. Soc. Amer. A*, vol. 10, no. 5, pp. 1014–23, May 1993.
[48] M. Makitalo and A. Foi, "Optimal inversion of the generalized anscombe transformation for Poisson-Gaussian noise," *IEEE Trans. Imag. Process.*, vol. 22, no. 1, pp. 91–103, Jan. 2013.
[49] G. Fatima and P. Babu, "PGPAL: A monotonic iterative algorithm for phase-retrieval under the presence of Poisson-Gaussian noise," *IEEE Signal Process. Lett.*, vol. 29, pp. 533–537, 2022.
[50] K. Bredies and M. Holler, "Regularization of linear inverse problems with total generalized variation," *J. Inverse Ill-Posed Problems*, vol. 22, no. 6, pp. 871–913, 2014.
[51] K. Bredies and T. Valkonen, "Inverse problems with second-order total generalized variation constraints," 2020, *arXiv:2005.09725*.
[52] I. Daubechies, Ten Lectures on Wavelets. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 1992. [Online]. Available: https://epubs.siam.org/doi/abs/10.1137/1.9781611970104
[53] G. Ongie, A. Jalal, C. A. Metzler, R. Baraniuk, A. G. Dimakis, and R. M. Willett, "Deep learning techniques for inverse problems in imaging," *IEEE J. Sel. Areas Inf. Theory*, vol. 1, no. 1, pp. 39–56, May 2020.
[54] A. Bora, A. Jalal, E. Price, and A. G. Dimakis, "Compressed sensing using generative models," in *Proc. 34th Int. Conf. Mach. Learn.*, 2017, vol. 70, pp. 537–546.
[55] S. H. Chan, X. Wang, and O. A. Elgendy, "Plug-and-play ADMM for image restoration: Fixed-point convergence and applications," *IEEE Trans. Comput. Imag.*, vol. 3, no. 1, pp. 84–98, Mar. 2017.
[56] K. Zhang, Y. Li, W. Zuo, L. Zhang, L. Van Gool, and R. Timofte, "Plug-and-play image restoration with deep denoiser prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 10, pp. 6360–6376, Oct. 2022.
[57] U. S. Kamilov, C. B. Bouman, G. T. Buzzard, and B. Wohlberg, "Plug-and-play methods for integrating physical and learned models in computational imaging," *IEEE Signal Process. Mag.*, vol. 40, no. 1, pp. 85–97, Jan. 2023.
[58] Y. Romano, M. Elad, and P. Milanfar, "The little engine that could: Regularization by denoising (red)," *SIAM J. Imag. Sci.*, vol. 10, no. 4, pp. 1804–1844, 2017.
[59] J. Lehtinen et al., "Noise2noise: Learning image restoration without clean data," in *Proc. 35th Int. Conf. Mach. Learn.*, 2018, pp. 2971–2980. [Online]. Available: http://proceedings.mlr.press/v80/lehtinen18a.html
[60] J. Batson and L. Royer, "Noise2Self: Blind denoising by self-supervision," in *Proc. 36th Int. Conf. Mach. Learn.*, 2019, vol. 97, pp. 524–533. [Online]. Available: https://proceedings.mlr.press/v97/batson19a.html

[61] Z. Wang, J. Liu, G. Li, and H. Han, "Blind2unblind: Self-supervised image denoising with visible blind spots," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 2017–2026.

[62] M. Asim, M. Daniels, O. Leong, A. Ahmed, and P. Hand, "Invertible generative models for inverse problems: Mitigating representation error and dataset bias," in *Proc. 37th Int. Conf. Mach. Learn.*, 2020, vol. 119, pp. 399–409. [Online]. Available: https://par.nsf.gov/biblio/10252230

[63] X. Wei, H. van Gorp, L. Gonzalez-Carabarin, D. Freedman, Y. C. Eldar, and R. J. G. van Sloun, "Deep unfolding with normalizing flow priors for inverse problems," *IEEE Trans. Signal Process.*, vol. 70, pp. 2962–2971, 2022.

[64] Y. Song and S. Ermon, "Generative modeling by estimating gradients of the data distribution," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, vol. 32. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2019/file/3001ef257407d5a371a96dcd947c7d93-Paper.pdf

[65] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Proc. 34th Int. Conf. Neural Inf. Process. Syst.*, 2020, vol. 33, pp. 6840–6851. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2020/file/4c5bcfec8584af0d967f1ab10179ca4b-Paper.pdf

[66] P. Dhariwal and A. Nichol, "Diffusion models beat gans on image synthesis," in *Proc. Adv. Neural Inf. Process. Syst.*, 2021, vol. 34, pp. 8780–8794. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2021/file/49ad23d1ec9fa4bd8d77d02681df5cfa-Paper.pdf

[67] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, "Score-based generative modeling through stochastic differential equations," in *Proc. Int. Conf. Learn. Representations*, 2021 [Online]. Available: https://openreview.net/forum?id=PxTIG12RRHS

[68] A. Graikos, N. Malkin, N. Jojic, and D. Samaras, "Diffusion models as plug-and-play priors," in *Proc. 36th Int. Conf. Neural Inf. Process. Syst.*, Red Hook, NY, USA, 2024, Art. no. 1070.

[69] S. Lee, H. Chung, J. Kim, and J. C. Ye, "Progressive deblurring of diffusion models for coarse-to-fine image synthesis," in *Proc. NeurIPS Workshop Score-Based Methods*, 2022. [Online]. Available: https://openreview.net/forum?id=KP8BrpZBbv

[70] A. Jalal, M. Arvinte, G. Daras, E. Price, A. G. Dimakis, and J. Tamir, "Robust compressed sensing MRI with deep generative priors," in *Proc. Adv. Neural Inf. Process. Syst.*, 2021, vol. 34, pp. 14938–14954. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2021/file/7d6044e95a16761171b130dcb476a43e-Paper.pdf

[71] H. Chung and J. C. Ye, "Score-based diffusion models for accelerated MRI," *Med. Image Anal.*, vol. 80, 2022, Art. no. 102479. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361841522001268

[72] Z.-X. Cui et al., "Self-score: Self-supervised learning on score-based models for MRI reconstruction," 2022, *arXiv:2209.00835*.

[73] Y. Long, J. A. Fessler, and J. M. Balter, "A 3D forward and back-projection method for X-ray CT using separable footprint," in *Proc. 1st Int. Mtg. Image Formation X-ray Computed Tomography*, 2009, pp. 146–9, winner of poster award. [Online]. Available: http://web.eecs.umich.edu/fessler/papers/files/proc/09/web/long-09-a3f.pdf

[74] Y. Song, L. Shen, L. Xing, and S. Ermon, "Solving inverse problems in medical imaging with score-based generative models," in *Proc. Int. Conf. Learn. Representations*, 2022. [Online]. Available: https://openreview.net/forum?id=vaRCHVj0uGI

[75] S. Shoushtari, J. Liu, and U. S. Kamilov, "Diffusion models for phase retrieval in computational imaging," in *Proc. IEEE Asilomar Conf. Signals, Syst, Comput.*, 2023, pp. 779–83.

[76] Z. Li, K. Lange, and J. A. Fessler, "Poisson phase retrieval with Wirtinger flow," in *2021 IEEE Int. Conf. Image Process.*, 2021, pp. 2828–2832.

[77] W. Luo, W. Alghamdi, and Y. M. Lu, "Optimal spectral initialization for signal recovery with applications to phase retrieval," *IEEE Trans. Signal Process.*, vol. 67, no. 9, pp. 2347–2356, May 2019.

[78] Z. Zhuang, D. Yang, F. Hofmann, D. Barmherzig, and J. Sun, "Practical phase retrieval using double deep image priors," in *Electronic Imaging*, 2022, pp 153-1–153-6, doi: 10.2352/EI.2023.35.14.COIMG-153.

[79] F. Wang et al., "Phase imaging with an untrained neural network," *Light, Sci. Appl.*, vol. 9, no. 1, pp. 77–77, 2020.

[80] Y. Zhang et al., "PhaseGAN: A deep-learning phase-retrieval approach for unpaired datasets," *Opt. Exp.*, vol. 29, no. 13, pp. 19593–19604, 2021. [Online]. Available: https://opg.optica.org/oe/abstract.cfm?URI=oe-29-13-19593

[81] D. A. Barmherzig and J. Sun, "Towards practical holographic coherent diffraction imaging via maximum likelihood estimation," *Opt. Exp.*, vol. 30, no. 5, pp. 6886–6906, 2022. [Online]. Available: https://opg.optica.org/oe/abstract.cfm?URI=oe-30-5-6886

[82] P. Thibault, M. Dierolf, A. Menzel, O. Bunk, C. David, and F. Pfeiffer, "High-resolution scanning X-ray diffraction microscopy," *Science*, vol. 321, no. 5887, pp. 379–382, 2008. [Online]. Available: https://www.science.org/doi/abs/10.1126/science.1158573

[83] S. Marchesini, A. Schirotzek, C. Yang, H. Wu, and F. Maia, "Augmented projections for ptychographic imaging," *Inverse Problems*, vol. 29, no. 11, 2013, Art. no. 115009, doi: 10.1088/0266-5611/29/11/115009.

[84] K. K. -W. Siu, A. Y. Nikulin, and P. Wells, "Unambiguous X-ray phase retrieval from fraunhofer diffraction data," *J. Appl. Phys.*, vol. 93, no. 5161, pp. 5161–5166, 2003.

[85] Y. Nesterov, "Smooth minimization of non-smooth functions," *Math. Prog.*, vol. 103, no. 1, pp. 127–152, 2005, doi: 10.1007/s10107-004-0552-5.

[86] D. Kim and J. A. Fessler, "Optimized first-order methods for smooth convex minimization," *Math Prog.*, vol. 159, no. 1, pp. 81–107, 2016.

[87] E. Bostan, M. Soltanolkotabi, D. Ren, and L. Waller, "Accelerated Wirtinger flow for multiplexed Fourier ptychographic microscopy," in *2018 25th IEEE Int. Conf. Image Process.*, 2018, pp. 3823–3827.

[88] Z. Fabian, J. Haldar, R. Leahy, and M. Soltanolkotabi, "3D phase retrieval at nano-scale via accelerated wirtinger flow," in *2020 28th Eur. Signal Process. Conf.*, 2021, pp. 2080–2084.

[89] Y. Gao, F. Yang, and L. Cao, "Pixel super-resolution phase retrieval for lensless on-chip microscopy via accelerated wirtinger flow," *Cells*, vol. 11, no. 13, Jun. 2022, doi: 10.3390/cells11131999.

[90] J. Fienup, "Phase retrieval algorithms: A comparison," *Appl. Opt.*, vol. 21, pp. 2758–2769, 1982.

[91] C. Metzler, P. Schniter, A. Veeraraghavan, and R. Baraniuk, "prDeep: Robust phase retrieval with a flexible deep network," in *Proc. 35th Int. Conf. Mach. Learn.*, 2018, vol. 80, pp. 3501–3510. [Online]. Available: https://proceedings.mlr.press/v80/metzler18a.html

[92] L.-H. Yeh et al., "Experimental robustness of Fourier ptychographic phase retrieval algorithms," *Opt. Express*, vol. 23, pp. 33214–33240, 2015.

[93] C. A. Metzler, A. Maleki, and R. G. Baraniuk, "BM3D-PRGAMP: Compressive phase retrieval based on BM3D denoising," in *2016 IEEE Int. Conf. Image Process.*, 2016, pp. 2504–2508.

[94] S. Sreehari et al., "Plug-and-play priors for bright field electron tomography and sparse interpolation," *IEEE Trans. Comput. Imag.*, vol. 2, no. 4, pp. 408–423, Dec. 2016.

[95] R. Ahmad et al., "Plug-and-play methods for magnetic resonance imaging: Using denoisers for image recovery," *IEEE Signal Process. Mag.*, vol. 37, no. 1, pp. 105–116, Jan. 2020.

[96] K. Zhang, W. Zuo, S. Gu, and L. Zhang, "Learning deep CNN denoiser prior for image restoration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 3929–3938.

[97] K. Zhang, W. Zuo, and L. Zhang, "Deep plug-and-play super-resolution for arbitrary blur kernels," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 1671–1681.

[98] K. Wei, A. Aviles-Rivero, J. Liang, Y. Fu, C.-B. Schönlieb, and H. Huang, "Tuning-free plug-and-play proximal algorithm for inverse imaging problems," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 10158–10169.

[99] K. Wei, A. Aviles-Rivero, J. Liang, Y. Fu, H. Huang, and C.-B. Schönlieb, "TFPnP: Tuning-free plug-and-play proximal algorithms with applications to inverse imaging problems," *J. Mach. Learn. Res.*, vol. 23, no. 16, pp. 1–48, 2022. [Online]. Available: http://jmlr.org/papers/v23/20-1297.html

[100] F. Heide, S. Diamond, M. Nießner, J. Ragan-Kelley, W. Heidrich, and G. Wetzstein, "Proximal: Efficient image optimization using proximal algorithms," *ACM Trans. Graph.*, vol. 35, no. 4, pp. 1–15, 2016, doi: 10.1145/2897824.2925875.

[101] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *Proc. IEEE Glob. Conf. Signal Process. Inf. Process.*, Dec. 2013, pp. 945–948.

[102] U. S. Kamilov, H. Mansour, and B. Wohlberg, "A plug-and-play priors approach for solving nonlinear imaging inverse problems," *IEEE Signal. Process. Lett.*, vol. 24, no. 12, pp. 1872–1876, Dec. 2017.

[103] Y. Romano, M. Elad, and P. Milanfar, "The little engine that could: Regularization by denoising (RED)," *SIAM J. Imag. Sci.*, vol. 10, no. 4, pp. 1804–1844, 2017.
[104] E. T. Reehorst and P. Schniter, "Regularization by denoising: Clarifications and new interpretations," *IEEE Trans. Comput. Imag.*, vol. 5, no. 1, pp. 52–67, Mar. 2019.
[105] V. Katkovnik and J. Astola, "Phase retrieval via spatial light modulator phase modulation in 4f optical setup: Numerical inverse imaging with sparse regularization for phase and amplitude," *J. Opt. Soc. Amer. A. Opt. Image Sci. Vis.*, no. 1, pp. 105–116, 2012.
[106] Y. Wang, X. Sun, and J. W. Fleischer, "When deep denoising meets iterative phase retrieval," in *Proc. 37th Int. Conf. Mach. Learn.*, 2020, Art. no. 928.
[107] P. Vincent, "A connection between score matching and denoising autoencoders," *Neural Computation*, vol. 23, no. 7, pp. 1661–1674, 2011.
[108] H. Chung, B. Sim, D. Ryu, and J. C. Ye, "Improving diffusion models for inverse problems using manifold constraints," in *Proc. Adv. Neural Inf. Process. Syst.*, 2022, vol. 35, pp. 25683–25696. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2022/file/a48e5877c7bf86a513950ab23b360498-Paper-Conference.pdf
[109] H. Chung, J. Kim, M. T. Mccann, M. L. Klasky, and J. C. Ye, "Diffusion posterior sampling for general noisy inverse problems," in *11th Int. Conf. Learn. Representations*, 2023. [Online]. Available: https://openreview.net/forum?id=OnD9zGAGT0k
[110] E. Chouzenoux, A. Jezierska, J.-C. Pesquet, and H. Talbot, "A convex approach for image restoration with exact Poisson–Gaussian likelihood," *SIAM J. Imag. Sci.*, vol. 8, no. 4, pp. 2662–2682, 2015.
[111] H. Li and Z. Lin, "Accelerated proximal gradient methods for nonconvex programming," in *Proc. Adv. Neural Inf. Process. Syst.*, C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, Eds., 2015, vol. 28. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2015/file/f7664060cc52bc6f3d620bcedc94a4b6-Paper.pdf
[112] A. Aksac et al., "BreCaHad: A dataset for breast cancer histopathological annotation and diagnosis," *BMC Res. Notes*, vol. 12, 2019, Art. no. 82.
[113] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proc. Int. Conf. Comput. Vis.*, 2015, pp. 3730–3738.
[114] Z. Li, J. A. Fessler, J. K. Mikell, S. J. Wilderman, and Y. K. Dewaraja, "Dblurdosenet: A deep residual learning network for voxel radionuclide dosimetry compensating for single-photon emission computerized tomography imaging resolution," *Med. Phys.*, vol. 49, no. 2, pp. 1216–1230, 2022.
[115] P. J. Huber, *Robust Statistics*. Hoboken, NJ, USA: Wiley, 1981.
[116] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.
[117] X. Xu, J. Liu, Y. Sun, B. Wohlberg, and U. S. Kamilov, "Boosting the performance of plug-and-play priors via denoiser scaling," in *Proc. 54th Asilomar Conf. Signals, Systems, Comput.*, 2020, pp. 1305–1312.
[118] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Representations*, San Diega, CA, USA, Dec. 2015.
[119] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[120] Z. Yuan and H. Wang, "Phase retrieval via reweighted wirtinger flow," *Appl. Opt.*, vol. 56, no. 9, pp. 2418–2427, 2017. [Online]. Available: https://opg.optica.org/ao/abstract.cfm?URI=ao-56-9-2418
[121] J.-F. Cai, M. Huang, D. Li, and Y. Wang, "Nearly optimal bounds for the global geometric landscape of phase retrieval," *Inverse Problems*, vol. 39, no. 7, 2023, Art. no. 075011, doi: 10.1088/1361-6420/acdab7.
[122] L. Grafakos, *Classical and Modern Fourier Analysis*. Hoboken, NJ, USA: Pearson/Prentice Hall, 2004.
[123] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016, [Online]. Available: http://www.deeplearningbook.org
[124] S. Ghadimi and G. Lan, "Accelerated gradient methods for nonconvex nonlinear and stochastic programming," *Math. Prog.*, vol. 156, pp. 59–99, 2016.
[125] Q. Ye, L.-W. Wang, and D. P. K. Lun, "Sisprnet: End-to-end learning for single-shot phase retrieval," *Opt. Exp.*, vol. 30, no. 18, pp. 31937–31958, 2022. [Online]. Available: https://opg.optica.org/oe/abstract.cfm?URI=oe-30-18-31937

SUPPLEMENT

This is the supplement for the paper "Accelerated Wirtinger Flow with Score-based Image Priors for Holographic Phase Retrieval in Poisson-Gaussian Noise Conditions" by Z. Li, J. Hu, X. Xu, L. Shen and J. A. Fessler.

A. *Reconstruction results of WF-Gaussian*

Fig. S.1 shows the reconstruction results using the WF-Gaussian method that assumes the variance of all measurements are the same, i.e., $\boldsymbol{y} = \mathcal{N}(|\boldsymbol{A}\boldsymbol{x}|^2 + \boldsymbol{b}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = c\boldsymbol{I}$. The data was simulated using the PG model, as described in Section III.

We tried both a line search method and the Fisher information approach for computing the step size, but neither resulted in a successful image recovery.

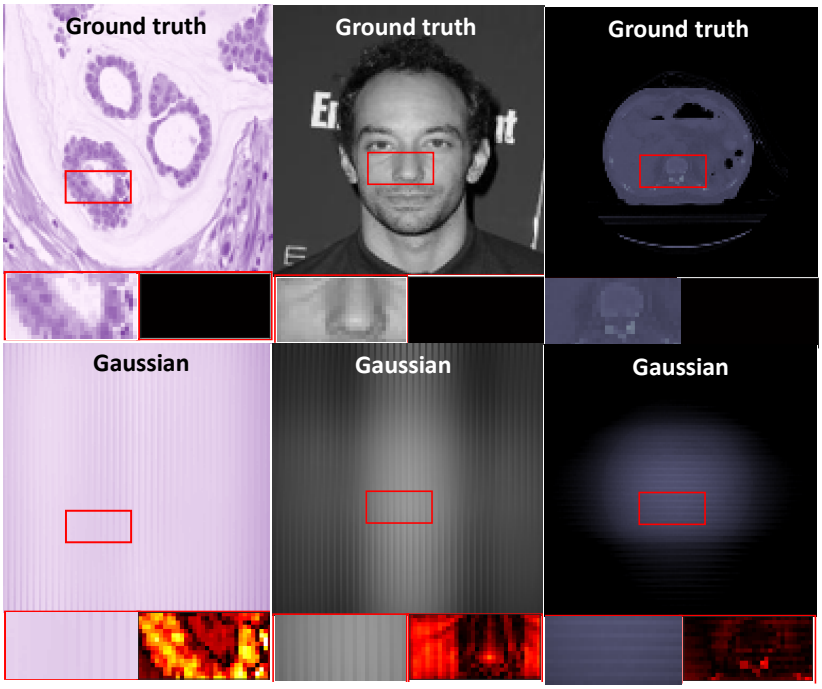


Fig. S.1: Reconstructed images using WF-Gaussian assuming the variance of all measurements are the same.

B. Proof of Theorem 2

It was already shown that ∇g_{PG} is Lipschitz continuous, so the remaining problem is to find a Lipschitz constant for $\boldsymbol{s}(\boldsymbol{x}, \sigma)$. We assume that the data allows the neural network to learn the score function well, i.e., $p_\sigma(\boldsymbol{x}) = p(\boldsymbol{x}) \circledast \mathcal{N}(0, \sigma^2)$, where $\circledast$ denotes (circular) convolution. We start with some well-known lemmas whose proofs can be readily found in [1]. We include the proofs for completeness.

Lemma 2.1: The Fourier transform (and inverse transform) of an absolutely integrable function is continuous (Exercise 2.2.6 in [1]).
Proof of Lemma 2.1: Let f be absolutely integrable and let $\tilde{f}$ be its Fourier transform. We have

$$|\tilde{f}(w+h) - \tilde{f}(w)| = \left| \int f(x)(e^{-2\pi j x(w+h)} - e^{-2\pi j w x}) \, \mathrm{d}x \right| \leq \int |f(x)||e^{2\pi j x h} - 1| \, \mathrm{d}x$$
$$\leq \max(|e^{2\pi j x h} - 1|) \int |f(x)| \, \mathrm{d}x \leq 2 \int |f(x)| \, \mathrm{d}x. \tag{S.1}$$

Using absolute integrability of f, we see $|\tilde{f}(w+h) - \tilde{f}(w)|$ tends to 0 as h tends to 0, so $\tilde{f}$ is uniformly continuous, which also implies it is continuous.

The proof of the inverse transform follows similarly.

Lemma 2.2: Suppose a sequence of functions $f_i : \mathbb{R} \to \mathbb{R}$ converges in $\mathcal{L}_1$ to some function f, and that each f_i is absolutely integrable. Then f is also absolutely integrable.
Proof: $\mathcal{L}_1$ consists of absolutely integrable functions and it is a Banach space so it is complete and hence closed. Thus, if f_i are a sequence in L1 then they are absolutely integrable. If this sequence converges, the limit must be in $\mathcal{L}_1$ because it is closed. This shows that f is absolutely integrable.

Proposition 2.1: The derivative of $\log(p_\sigma(x))$ is bounded on the interval $[-C, C]$.
Proof: We start by dropping constant factors and using derivative of a convolution, we have

$$\frac{d}{dx}(\log(p(x) \circledast \mathcal{N}(0, \sigma^2))) \sim \frac{\mathcal{F}^{-1}(\imath x \mathcal{F}(p(x)) \cdot \mathcal{F}(\mathcal{N}(0, \sigma^2)))}{p(x) \circledast \mathcal{N}(0, \sigma^2)} \sim \frac{\mathcal{F}^{-1}(x e^{-x^2} \cdot \mathcal{F}(p(x)))}{p(x) \circledast \mathcal{N}(0, \sigma^2)} \tag{S.2}$$

where $\mathcal{F}$ denotes Fourier transform. The denominator is continuous and since x lies in a closed interval by assumption, has a lower bound $M > 0$ by the extreme value theorem. We next consider the numerator.
By [2, pp. 65], a sequence of Gaussian mixture models (GMMs) can be used to approximate any smooth probability distribution in $\mathcal{L}_2$ convergence. Furthermore, $\mathcal{L}_2$ convergence implies $\mathcal{L}_1$ convergence. Hence, consider a sequence of GMMs f_i that converge in $\mathcal{L}_1$ to $p(x)$. By linearity of Fourier transform, $\mathcal{F}(f_i(x))$ must be a linear combination of terms of the form $e^{-(x-\mu_i)^2/c_i}$ for some c_i. Thus, the numerator $xe^{-x^2} \cdot \mathcal{F}(f_i(x))$ is a finite linear combination of terms of the form $xe^{-(x-\mu_i)^2/c_i}$, each of which are absolutely integrable. Therefore, we have a sequence of functions, each of which are absolutely integrable, that converge in $\mathcal{L}_1$ to $xe^{-x^2} \cdot \mathcal{F}(p(x))$, so by Lemma 2.2, this is also absolutely integrable. By Lemma 2.1, the inverse Fourier transform of this is continuous. Finally, again by the extreme value theorem and using the boundedness of x, the numerator is bounded above by some $M' > 0$. Hence, the entire expression (S.2) is bounded above by M'/M.

Lemma 2.4: Suppose we have an everywhere twice differentiable function of two variables $f(x, y) : \mathbb{R}^2 \to \mathbb{R}$. Then $\frac{\partial^2}{\partial x \partial y} \log f(x, y)$ is bounded if the following three conditions are met:

1) $\frac{\partial^2}{\partial x \partial y} f(x, y)$ is bounded.
2) f itself is bounded below by a positive number and also bounded above.
3) ∇f is bounded.

Proof: Suppose we have $f(x, y)$ satisfying those three conditions. We compute the second partial derivative of its log:

$$\frac{\partial}{\partial x} \log f(x, y) = \frac{\frac{\partial}{\partial x} f(x, y)}{f(x, y)}. \tag{S.3}$$

and

$$\frac{\partial^2}{\partial x \partial y} \log f(x, y) = \frac{(\frac{\partial^2}{\partial x \partial y} f(x, y)) f(x, y) - (\frac{\partial}{\partial x} f(x, y))(\frac{\partial}{\partial y} f(x, y))}{f(x, y)^2}. \tag{S.4}$$

From the second condition, the denominator is bounded below by a positive number, so it suffices to consider the boundedness

of the numerator. The first term of the numerator is a product of two quantities, the first of which is bounded by the first condition and the second of which is bounded by the second condition. The second term of the numerator is also a product of two quantities, both of which are bounded by the third condition. Thus, this shows $\frac{\partial^2}{\partial x \partial y} \log f(x,y)$ is bounded. □

Proposition 2.2: The gradient of $\log(p_\sigma(\boldsymbol{x}))$ is Lipschitz continuous on $[-C,C]^N$.
Proof: By renaming the variables, and redefining $f(x,y) = p(x,y,\cdots)$, we may consider the boundedness of $\frac{\partial^2}{\partial x \partial y}\left(\log f(x,y) \circledast \mathcal{N}(0,\sigma^2\boldsymbol{I})\right)$ on $[-C,C]^2$. To apply Lemma 2.4 to remove the log, we need to verify the three conditions. Define $g(x,y) = f(x,y) \circledast \mathcal{N}(0,\sigma^2\boldsymbol{I})$. The second condition is readily verified to be true: By assumption, x and y take values on a closed interval, thus by the extreme value theorem, so does $g(x,y)$. Further, g is a convolution of positive numbers and so the output is always positive, hence, the lower bound of this closed interval is a positive number, verifying this condition.

For the third condition, we need to consider boundedness of $\frac{\partial}{\partial x} f(x,y) \circledast \mathcal{N}(0,\sigma^2\boldsymbol{I})$. This is nearly identical to Proposition 2.2, with the only difference being we have some general function in terms of only x $f(x,y)$ instead of a probability distribution $p(x)$. The proof of that lemma is readily adapted for this case with the only condition needing verification being the absolute integrability over x of $f(x,y)$. In fact, this is clear because f is always positive; hence, integrating over f with respect to x must yield a finite number as integrating a second time over y yields 1.

It thus suffices to consider boundedness of $h(x,y) = \frac{\partial^2}{\partial x \partial y}\left(f(x,y) \circledast \mathcal{N}(0,\sigma^2\boldsymbol{I})\right)$. It is assumed that f is smooth and the convolution of smooth functions is smooth, which implies $f(x,y) \circledast \mathcal{N}(0,\sigma^2\boldsymbol{I})$ is smooth. Hence h is differentiable, so it is continuous. Once again by the EVT, as x and y take values on a closed interval, h must be bounded. By Lemma 2.4, the entries of the Hessian of the score function are bounded. Therefore, a Lipschitz constant of $s_{\boldsymbol{\theta}}(\boldsymbol{x})$ exists. □

Proof of Theorem 2: By Proposition 2.2, and from the design of Algorithm 2 (see main text), $\boldsymbol{x}_{t,k}$ and $\boldsymbol{w}_{t,k}$ are both bounded between $[0,C]$ for all t,k, so the Lipschitz constant $\mathcal{L}^*$ of $\nabla g_{\mathrm{PG}}(\cdot) + s_{\boldsymbol{\theta}}(\cdot)$ exists. With the step size μ satisfying $0 < \mu < \frac{1}{\mathcal{L}^*}$, and the weighting factor $\boldsymbol{\gamma} \in \{0,1\}$ being chosen according to whichever higher posterior probability between $p_{\sigma_k}(\boldsymbol{z}|\boldsymbol{A},\boldsymbol{y},\bar{\boldsymbol{b}},\boldsymbol{r})$ and $p_{\sigma_k}(\boldsymbol{v}|\boldsymbol{A},\boldsymbol{y},\bar{\boldsymbol{b}},\boldsymbol{r})$ (where $p_{\sigma_k}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{A},\bar{\boldsymbol{b}},\boldsymbol{r}) \propto p(\boldsymbol{y}|\boldsymbol{A},\boldsymbol{x},\bar{\boldsymbol{b}},\boldsymbol{r}) p_{\sigma_k}(\boldsymbol{x})$), then we satisfy all conditions in Theorem 1 of [3], which establishes the critical-point convergence of the inner iteration sequence $\boldsymbol{x}_{t,k}$ in Algorithm 2 (see main text) for the posterior distribution $p_{\sigma_k}(\boldsymbol{x}|\boldsymbol{A},\boldsymbol{y},\bar{\boldsymbol{b}},\boldsymbol{r})$. Similar convergence analysis can be found in [4]. □

Algorithm S.1 Poisson-Gaussian phase retrieval via WFSD.

Require: Measurement $\boldsymbol{y}$, system matrix $\boldsymbol{A}$, step size factor ϵ, truncation operator $\mathcal{P}_C(\cdot) \to [0, C]$; initial image $\boldsymbol{x}_0$, initialize $\sigma_1 > \sigma_2 > \cdots > \sigma_K$.
for $k = 1 : K$ **do**
  for $t = 1 : T$ **do**
    Set step size $\mu = \epsilon\sigma_k^2$.
    Compute $s_{\boldsymbol{\theta}}(\boldsymbol{x}_{t,k}, \sigma_k)$.
    Set $\boldsymbol{x}_{t+1,k} = \mathcal{P}_C\left(\boldsymbol{x}_{t,k} - \mu\left(\nabla g_{\mathrm{PG}}(\boldsymbol{x}_{t,k}) + s_{\boldsymbol{\theta}}(\boldsymbol{x}_{t,k}, \sigma_k)\right)\right)$.
  end for
end for
Return $\boldsymbol{x}_{T,K}$.

C. Algorithm Implementation

Wirtinger Flow. WF is a popular algorithm for phase retrieval. It first computes the Wirtinger gradient (an ascent direction) and then applies gradient descent. Perhaps the most critical step is to find an appropriate step size. In this work, we used backtracking line search for Gaussian WF, and the observed Fisher information [5] for Poisson WF and for Poisson-Gaussian WF. To further accelerate the computation (and avoid floating point overflow) of ∇g_{PG}, we observed it was effective to use the gradient of Poisson PR for large y_i, e.g., $y_i \geq 100$. We used this "trick" in our experiments; additionally, one can also use "defocus" to deal with large y_i [6]. By replacing $\nabla h(\boldsymbol{x})$ in Algorithm 1 (see main text) with the trained score function $s_{\boldsymbol{\theta}}$, one can derive the vanilla gradient-descent version (Algorithm S.1) of the AWFSD algorithm.

PnP-ADMM. The plug-and-play ADMM first derives a Lagrangian using variable splitting and then applies alternating minimization [7]. In this work, let $\boldsymbol{u} = \boldsymbol{x}$, and the Lagrangian is

$$L(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\eta}; \rho) = g_{\mathrm{PG}}(|\boldsymbol{A}\boldsymbol{u}|^2 + \boldsymbol{b}) + R(\boldsymbol{x}) + \frac{\rho}{2}\left(\|\boldsymbol{x} - \boldsymbol{u} + \boldsymbol{\eta}\|_2^2 - \|\boldsymbol{\eta}\|_2^2\right). \tag{S.5}$$

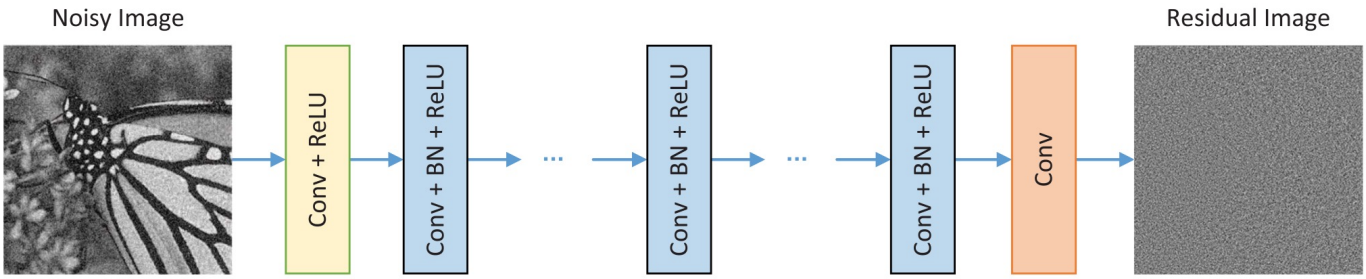


Fig. S.2: *The architecture of the adopted DnCNN network [8].*

DOLPH. Algorithm A.2 shows our implementation of DOLPH, which is based on the DDPM model [9] that first gradually adds Gaussian noise to data according to a variance schedule $\beta_1, \cdots, \beta_T$ so that $q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) = \mathcal{N}(\boldsymbol{x}_t; \sqrt{1-\beta_t}\boldsymbol{x}_{t-1}, \beta_t\boldsymbol{I})$. Using the notation $\alpha_t = 1 - \beta_t$ and $\overline{\alpha_t} = \prod_{s=1}^{t} \alpha_s$, we have

$$q(\boldsymbol{x}_t|\boldsymbol{x}_0) = \mathcal{N}(\boldsymbol{x}_t; \sqrt{\overline{\alpha_t}}\boldsymbol{x}_0, (1 - \overline{\alpha_t})\boldsymbol{I}). \tag{S.6}$$

It can be shown [9] that the appropriate loss function to use is

$$L(\theta) = \mathbb{E}_{t, \boldsymbol{x}_0, \epsilon}\left[\|\epsilon - \epsilon_\theta(\sqrt{\overline{\alpha_t}}\boldsymbol{x}_0 + \sqrt{1-\overline{\alpha_t}}\epsilon, t)\|^2\right], \tag{S.7}$$

where ϵ is selected from $\mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$. The sampling and reconstruction algorithm requires the addition of noise each step as shown in the following algorithm. Experimentally, however, we found that setting $\sigma_t = 0$ results in higher quality reconstructed images. Furthermore, we choose $T = 100$, $\beta_1 = 10^{-4}$, and $\beta_T = 0.3$. For the theory to hold, $\boldsymbol{x}_T$ should be indistinguishable from white Gaussian noise, which is readily verified to be true for these parameters. Finally, the stepsize μ_k of the gradient descent step can be chosen according to the Lipschitz constant of the Poisson-Gaussian likelihood to ensure convergence, or empirically, as is done in the experiments.

Algorithm S.3 summarizes the PnP-ADMM algorithm for phase retrieval. In this work, we trained the denoiser $h_{\boldsymbol{\theta}}$ using the network DnCNN [8]. As shown in Fig. S.2, the architecture of DnCNN consists of convolution (Conv), Rectified Linear Unit (ReLU), and batch normalization(BN). The network was trained with residual learning where the output is the noise residual

and the clean image was obtained by subtracting the noisy output. We trained all denoisers for 400 epochs with image patches of size 40×40 on each given dataset.

Algorithm S.2 DOLPH [10].

Require: Measurement $\boldsymbol{y}$, system matrix $\boldsymbol{A}$, initialization of image $\boldsymbol{x}_T$, pre-trained DDPM model $s_{\boldsymbol{\theta}}$, and T.
  for $t = T : 1$ **do**
    Set $\boldsymbol{z}_t \sim \mathcal{N}(0, \boldsymbol{I})$ if $t > 1$ and $\boldsymbol{z}_t = 0$ otherwise.
    Set $\boldsymbol{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(\boldsymbol{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}} \boldsymbol{s}_\theta(\boldsymbol{x}_t, t) \right) + \sigma_t \boldsymbol{z}_t$.
    Determine step size μ.
    Set $\boldsymbol{x}_{t-1} = \boldsymbol{x}_{t-1} - \mu \nabla g_{\mathrm{PG}}(\boldsymbol{x}_{t-1})$.
  end for
Return $\boldsymbol{x}_1$.

Algorithm S.3 Poisson-Gaussian phase retrieval via PnP-ADMM

Require: Measurement $\boldsymbol{y}$, system matrix $\boldsymbol{A}$, initialization of image $\boldsymbol{x}_0$, initialization of auxilary variable $\boldsymbol{u}_0 = \boldsymbol{x}_0$, initialization of dual variable $\boldsymbol{\eta}_0 = \boldsymbol{0}$, pre-trained denoiser $h_{\boldsymbol{\theta}}$, Lagrangian penalty parameter ρ.
  for $k = 1 : K$ **do**
    for $t = 1 : T$ **do**
      Compute step size $\mu_{t,k}$.
      Set $\boldsymbol{u}_{t+1,k} = \boldsymbol{u}_{t,k} - \mu_{t,k} \left(\nabla g_{\mathrm{PG}}(\boldsymbol{u}_{k,t}) + \rho(\boldsymbol{u}_{t,k} - \boldsymbol{x}_k - \boldsymbol{\eta}_k) \right)$.
    end for
    Set $\boldsymbol{x}_{k+1} = h_{\boldsymbol{\theta}}(\boldsymbol{x}_k)$.
    Set $\boldsymbol{\eta}_{k+1} = \boldsymbol{\eta}_k + \boldsymbol{x}_{k+1} - \boldsymbol{u}_{k+1}$.
  end for
Return $\boldsymbol{x}_K$.

PnP-PGM. Similar to PnP-ADMM, one can also derive a proximal gradient method as shown in Algorithm S.4. Here we assume the denoising of $h_{\boldsymbol{\theta}}$ is a proximal operation.

Algorithm S.4 Poisson-Gaussian phase retrieval via PnP-PGM

Require: Measurement $\boldsymbol{y}$, system matrix $\boldsymbol{A}$, initialization of image $\boldsymbol{x}_0$, pre-trained denoiser $h_{\boldsymbol{\theta}}$, averaging factor β.
  for $k = 1 : K$ **do**
    Compute step size μ_k.
    Set $\tilde{\boldsymbol{x}}_k = \boldsymbol{x}_k - \mu_k \nabla g_{\mathrm{PG}}(\boldsymbol{x}_k)$.
    Set $\bar{\boldsymbol{x}}_k = h_{\boldsymbol{\theta}}(\tilde{\boldsymbol{x}}_k)$.
    Set $\boldsymbol{x}_{k+1} = (1 - \beta)\tilde{\boldsymbol{x}}_k + \beta \bar{\boldsymbol{x}}_k$.
  end for
Return $\boldsymbol{x}_K$.

SD-RED. Regularization by denoising (RED) is an alternative to PnP methods that is based on an explicit image-adaptive regularization functional: $\frac{1}{2}\boldsymbol{x}' \left(\boldsymbol{x} - h_{\boldsymbol{\theta}}(\boldsymbol{x}) \right)$. This regularizer reflects the cross-correlation between the image and its denoising residual [11].

Algorithm S.5 summarizes the RED approach for phase retrieval.

Algorithm S.5 Poisson-Gaussian phase retrieval via RED

Require: Measurement $\boldsymbol{y}$, system matrix $\boldsymbol{A}$, initialization of image $\boldsymbol{x}_0$, pre-trained denoiser $h_{\boldsymbol{\theta}}$, regularization factor β.
  for $k = 1 : K$ **do**
    Compute stepsize μ_k.
    Set $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \mu_k \left(\nabla g_{\mathrm{PG}}(\boldsymbol{x}_k) + \beta(\boldsymbol{x}_k - h_{\boldsymbol{\theta}}(\boldsymbol{x}_k)) \right)$.
  end for
Return $\boldsymbol{x}_K$.

SD-RED-SELF. Other than supervised denoising approaches, we also implemented a self-supervised denoising method known as "noise2self" [12], which designed a neural network to be $\mathcal{J}$-invariant so that the self-supervised loss can be represented as the sum of supervised loss and the variance of noise. The "SD-RED-SELF" algorithm refers to training $h_{\boldsymbol{\theta}}$ in Algorithm S.5 in such self-supervised fashion on each test data.

REFERENCES

[1] L. Grafakos. *Classical and Modern Fourier Analysis*. Pearson/Prentice Hall, 2004.

[2] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. http://www.deeplearningbook.org. MIT Press, 2016.

[3] H. Li and Z. Lin. "Accelerated Proximal Gradient Methods for Nonconvex Programming". In: *Advances in Neural Information Processing Systems*. Vol. 28. 2015.

[4] S. Ghadimi and G. Lan. "Accelerated gradient methods for nonconvex nonlinear and stochastic programming". In: *Math. Program.* 156 (2016). DOI: https://doi.org/10.1007/s10107-015-0871-8.

[5] Z. Li, K. Lange, and J. A. Fessler. "Poisson Phase Retrieval in Very Low-Count Regimes". In: *IEEE Transactions on Computational Imaging* 8 (2022), pp. 838–850. DOI: 10.1109/TCI.2022.3209936.

[6] Q. Ye, L.-W. Wang, and D. P. K. Lun. "SiSPRNet: end-to-end learning for single-shot phase retrieval". In: *Opt. Express* 30.18 (2022), pp. 31937–31958. DOI: 10.1364/OE.464086.

[7] S. H. Chan, X. Wang, and O. A. Elgendy. "Plug-and-Play ADMM for Image Restoration: Fixed-Point Convergence and Applications". In: *IEEE Transactions on Computational Imaging* 3.1 (2017), pp. 84–98. DOI: 10.1109/TCI.2016.2629286.

[8] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang. "Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising". In: *IEEE Transactions on Image Processing* 26.7 (2017), pp. 3142–3155. DOI: 10.1109/TIP.2017.2662206.

[9] J. Ho, A. Jain, and P. Abbeel. "Denoising Diffusion Probabilistic Models". In: 33 (2020). Ed. by H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, pp. 6840–6851.

[10] S. Shoushtari, J. Liu, and U. S. Kamilov. "Diffusion models for phase retrieval in computational imaging". In: *Proc., IEEE Asilomar Conf. on Signals, Systems, and Comp.* 2023, 779–83. DOI: 10.1109/IEEECONF59524.2023.10477083.

[11] Y. Romano, M. Elad, and P. Milanfar. "The Little Engine That Could: Regularization by Denoising (RED)". In: *SIAM Journal on Imaging Sciences* 10.4 (2017), pp. 1804–1844. DOI: 10.1137/16M1102884.

[12] J. Batson and L. Royer. "Noise2Self: Blind Denoising by Self-Supervision". In: *Proceedings of the 36th International Conference on Machine Learning*. Ed. by K. Chaudhuri and R. Salakhutdinov. Vol. 97. Proceedings of Machine Learning Research. PMLR, June 2019, pp. 524–533.