Complex-Valued Retrievals From Noisy Images Using Diffusion Models
Supplementary Material

Nadav Torem[*†]   Roi Ronen[*†]   Yoav Y. Schechner[†]   Michael Elad[‡]

[†]Viterbi Faculty of Electrical and Computer Engineering & [‡]Department of Computer Science
Technion - Israel Institute of Technology, Haifa, Israel

{torem@campus, roironen@campus, yoav@ee, elad@cs}.technion.ac.il

Abstract

This is a supplementary document to the main manuscript. Here we provide more quantitative and qualitative results. Moreover, this document details gradient derivations, which were not included in the main manuscript, for space limits.

1. Denoising Poissonian Image Intensities

This section details the complete derivations in Sec. 3.1 of the main manuscript. To be self-contained, this section repeats some equations that already exist in the main manuscript. In this section, all vector operations are *element-wise*. Denote incorporation of noise into a clean image $\mathbf{x}$ by the operator $\mathcal{N}$. The noisy measurement is

$$\mathbf{y} = \mathcal{N}(\mathbf{x}) \ . \qquad (1)$$

Let $N(0, \sigma^2)$ be a Gaussian distribution with zero expectation and variance σ^2. Let σ_0 be a noise level, known from the camera properties, as described in Sec. 2.2 of the main manuscript. We set a sequence of noise levels

$$\sigma_0 > \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_T > \sigma_{T+1} = 0 \ . \qquad (2)$$

For analysis, we act as if $\mathbf{y}$ is related by Eq. (5) of the main manuscript,

$$\mathbf{y} = \mathbf{x} + \mathbf{n}^{\text{meas}} \ , \qquad (3)$$

where

$$\mathbf{n}^{\text{meas}} \sim N(0, \sigma_0^2 \mathbf{x}) \ . \qquad (4)$$

Recall from the main manuscript that $t \in [1, \dots, T]$ is an iteration index, $\alpha_t > 0$ is an appropriately chosen small function of t, and $\mathbf{I}$ is the identity matrix. Then, annealed Langevin dynamics has the following iterative rule

$$\tilde{\mathbf{x}}_{t+1} = \tilde{\mathbf{x}}_t + \alpha_t \nabla_{\tilde{\mathbf{x}}_t} \log p(\tilde{\mathbf{x}}_t|\mathbf{y}) + \sqrt{2\alpha_t}\mathbf{n}_t^{\text{Langevin}} \ . \qquad (5)$$

Using Bayes rule,

$$\nabla_{\tilde{\mathbf{x}}_t} \log p(\tilde{\mathbf{x}}_t|\mathbf{y}) = \nabla_{\tilde{\mathbf{x}}_t} \left[\log p(\mathbf{y}|\tilde{\mathbf{x}}_t) + \log p(\tilde{\mathbf{x}}_t)\right] \ . \qquad (6)$$

Here, $\mathbf{n}_t^{\text{Langevin}} \sim N(0, \mathbf{I})$ is a *synthetic* noise at iteration t.

In Section 1.1 herein, we detail $\nabla_{\tilde{\mathbf{x}}_t} \log p(\mathbf{y}|\tilde{\mathbf{x}}_t)$, while $\nabla_{\tilde{\mathbf{x}}_t} \log p(\tilde{\mathbf{x}}_t)$ is described in Section 1.2 herein. Overall, the Langevin dynamics algorithm for denoising Poissonian image intensities is shown in Algorithm 1 herein. The small constant ϵ in Algorithm 1 herein is a pre-defined step size.

1.1. Likelihood Gradient: Poisson Denoising

Let $\mathbf{e}_t^{\text{part}}$ be a statistically independent synthetic *partial* dummy error term, distributed as

$$\mathbf{e}_t^{\text{part}} \sim N\left[0, (\sigma_t^2 - \sigma_{t+1}^2)\mathbf{I}\right] \ . \qquad (7)$$

A sum of zero-mean Gaussian variables is a Gaussian, whose variance is the sum of the variances. So,

$$\sum_{\tau=t_1}^{t_2} \mathbf{e}_\tau^{\text{part}} \sim$$
$$N\left[0, \sum_{\tau=t_1}^{t_2}(\sigma_\tau^2 - \sigma_{\tau+1}^2)\mathbf{I}\right] = N\left[0, (\sigma_{t_1}^2 - \sigma_{t_2}^2)\mathbf{I}\right] \ . \qquad (8)$$

Note that using Eqs. (2) and (8) herein

$$\sum_{\tau=0}^{T} \mathbf{e}_\tau^{\text{part}} \sim N\left(0, \sigma_0^2\mathbf{I}\right) \ . \qquad (9)$$

Using Eqs. (4) and (9) herein, we design $\mathbf{e}^{\text{part}}$ such that

$$\mathbf{n}^{\text{meas}} = N(0, \sigma_0^2\mathbf{x}) = \sqrt{\mathbf{x}}N(0, \sigma_0^2\mathbf{I}) = \sqrt{\mathbf{x}}\sum_{\tau=0}^{T} \mathbf{e}_\tau^{\text{part}} \ . \qquad (10)$$

Let $\mathbf{e}_t^{\text{annea}}$ be a dummy error term. We design $\mathbf{e}_t^{\text{annea}}$ such that it is conceptually contained in the measurement

Algorithm 1: Langevin dynamics for denoising Poissonian image intensities

---

Input: $\{\sigma_t\}_{t=1}^T, \mathbf{y}, \sigma_0$ and a small constant ϵ
Initialize $\tilde{\mathbf{x}}_0 \sim N(0, \sigma_0^2 \mathbf{I})$
for $t \leftarrow 1$ *to* T **do**
    $\alpha_t \leftarrow \epsilon \frac{\sigma_t^2}{\sigma_T^2}$
    Draw $\mathbf{n}_t^{\text{Langevin}} \sim N(0, \mathbf{I})$
    $\boldsymbol{\Delta}_t \leftarrow \nabla_{\tilde{\mathbf{x}}_{t-1}} \log p(\mathbf{y}|\tilde{\mathbf{x}}_{t-1}) + h_{\hat{\Theta}}^{\text{Poisson}}(\tilde{\mathbf{x}}_{t-1}, \sigma_t)$
    $\tilde{\mathbf{x}}_t \leftarrow \tilde{\mathbf{x}}_{t-1} + \alpha_t \boldsymbol{\Delta}_t + \sqrt{2\alpha_t}\mathbf{n}_t^{\text{Langevin}}$
end
$\hat{\mathbf{x}} \leftarrow \tilde{\mathbf{x}}_T$

---

noise. This enables to derive an analytical expression for $\nabla_{\tilde{\mathbf{x}}_t} \log p(\mathbf{y}|\tilde{\mathbf{x}}_t)$. We set

$$\mathbf{e}_t^{\text{annea}} = \sum_{\tau=t}^T \mathbf{e}_\tau^{\text{part}} . \tag{11}$$

We define a relation between the unknown desired variable $\mathbf{x}$ and the annealed variable as

$$\tilde{\mathbf{x}}_t = \mathbf{x} + \sqrt{\mathbf{x}}\mathbf{e}_t^{\text{annea}}. \tag{12}$$

Inserting Eqs. (10) and (12) herein in Eq. (3) herein yields

$$\mathbf{y} = \tilde{\mathbf{x}}_t - \sqrt{\mathbf{x}}\sum_{\tau=t}^T \mathbf{e}_\tau^{\text{part}} + \sqrt{\mathbf{x}}\sum_{\tau=0}^T \mathbf{e}_\tau^{\text{part}} = \tilde{\mathbf{x}}_t + \sqrt{\mathbf{x}}\sum_{\tau=0}^{t-1} \mathbf{e}_\tau^{\text{part}}. \tag{13}$$

To derive $\nabla_{\tilde{\mathbf{x}}_t} \log p(\mathbf{y}|\tilde{\mathbf{x}}_t)$, we seek to express $\mathbf{y}$ using terms (besides $\tilde{\mathbf{x}}_t$) whose probability distributions are known and independent of $\tilde{\mathbf{x}}_t$. However, in Eq. (13) herein, $\mathbf{x}$ is unknown. We can hypothetically extract $\mathbf{x}$ from Eq. (12) herein using

$$\mathbf{x} = \frac{1}{2}\left[2\tilde{\mathbf{x}}_t \pm \sqrt{4\tilde{\mathbf{x}}_t + (\mathbf{e}_t^{\text{annea}})^2}\mathbf{e}_t^{\text{annea}} + (\mathbf{e}_t^{\text{annea}})^2\right], \tag{14}$$

and substitute Eq. (14) herein in Eq. (13) herein. This would make $\mathbf{y}$ expressed using three terms: $\tilde{\mathbf{x}}_t, \mathbf{e}_t^{\text{annea}}$ and $\sum_{\tau=0}^{t-1} \mathbf{e}_\tau^{\text{part}}$. However, from Eq. (12) herein, $\tilde{\mathbf{x}}_t, \mathbf{e}_t^{\text{annea}}$ *are statistically dependent*. Thus, we cannot use Eq. (14).

Instead, we apply the following steps. From Eqs. (8) and (13) herein,

$$p(\mathbf{y}|\tilde{\mathbf{x}}_t) = p(\mathbf{y} - \tilde{\mathbf{x}}_t|\tilde{\mathbf{x}}_t) = p\left(\sqrt{\mathbf{x}}\sum_{\tau=0}^{t-1}\mathbf{e}_\tau^{\text{part}}|\tilde{\mathbf{x}}_t\right) \tag{15}$$

Recall that $\mathbf{x}$ in Eq. (15) herein is unknown. Therefore, to set an explicit analytical expression of Eq. (15), we approximate the remaining measurement noise as if the signal

dependency is proportional to $\tilde{\mathbf{x}}_t$ rather then $\mathbf{x}$[1]. Overall, Eq. (15) takes the form of

$$p\left(\sqrt{\tilde{\mathbf{x}}_t}\sum_{\tau=0}^{t-1}\mathbf{e}_\tau^{\text{part}}|\tilde{\mathbf{x}}_t\right). \tag{16}$$

Finally, the expression $\sqrt{\tilde{\mathbf{x}}_t}\sum_{\tau=0}^{t-1}\mathbf{e}_\tau^{\text{part}}$ is both known and statistically independent of $\tilde{\mathbf{x}}_t$. This allows us to find an explicit expression of the probability distribution $p(\mathbf{y}|\tilde{\mathbf{x}}_t)$

$$p(\mathbf{y}|\tilde{\mathbf{x}}_t) \approx \frac{1}{\sqrt{2\pi(\sigma_0^2 - \sigma_t^2)\tilde{\mathbf{x}}_t}}\exp\left[-\frac{(\mathbf{y} - \tilde{\mathbf{x}}_t)^2}{2(\sigma_0^2 - \sigma_t^2)\tilde{\mathbf{x}}_t}\right]. \tag{17}$$

Computing the gradient of the logarithm of Eq. (17) herein results in

$$\nabla_{\tilde{\mathbf{x}}_t} \log p(\mathbf{y}|\tilde{\mathbf{x}}_t) = \frac{1}{2(\sigma_0^2 - \sigma_t^2)}\left(\frac{\mathbf{y}^2}{\tilde{\mathbf{x}}_t^2} - 1\right) - \frac{1}{2\tilde{\mathbf{x}}_t}. \tag{18}$$

1.2. Gradient of the Prior: Poisson Denoising

We need to derive $\nabla_{\tilde{\mathbf{x}}_t} \log p(\tilde{\mathbf{x}}_t)$ in Eq. (6) herein for iterations of annealed Langevin dynamics. So, before imaging takes place, we pre-train a deep neural network (DNN) $h_\Theta^{\text{Poisson}}$, to approximate $\nabla_{\tilde{\mathbf{x}}_t} \log p(\tilde{\mathbf{x}}_t)$. In analogy to Eq. (12) herein, let $(\mathbf{x}, \mathbf{x}')$ be a synthetic pair of clean and noisy images, rendered using

$$\mathbf{x}' = \mathbf{x} + \sqrt{\mathbf{x}}\mathbf{n}^{\text{train}}. \tag{19}$$

Here, $\mathbf{n}^{\text{train}} \sim N(0, \sigma^2\mathbf{I})$, where σ is an arbitrary noise level. Using Eq. (19) herein,

$$p(\mathbf{x}'|\mathbf{x}) = p(\mathbf{x}' - \mathbf{x}|\mathbf{x}) = p\left(\sqrt{\mathbf{x}}\mathbf{n}^{\text{train}}|\mathbf{x}\right)$$
$$= \frac{1}{\sqrt{2\pi\sigma^2\mathbf{x}}}\exp\left[-\frac{(\mathbf{x}' - \mathbf{x})^2}{2\sigma^2\mathbf{x}}\right]. \tag{20}$$

By definition, the marginal distribution $p(\mathbf{x}')$ can be written as,

$$p(\mathbf{x}') = \int p(\mathbf{x}'|\mathbf{x})p(\mathbf{x})d\mathbf{x}. \tag{21}$$

Inserting Eq. (20) in Eq. (21) herein yields

$$p(\mathbf{x}') = \int \frac{1}{\sqrt{2\pi\sigma^2\mathbf{x}}}\exp\left[-\frac{(\mathbf{x}' - \mathbf{x})^2}{2\sigma^2\mathbf{x}}\right]p(\mathbf{x})d\mathbf{x}. \tag{22}$$

Then, the gradient of Eq. (22) herein satisfies

$$\nabla_{\mathbf{x}'}p(\mathbf{x}') = \int \frac{1}{\sqrt{2\pi\sigma^2\mathbf{x}}}\nabla_{\mathbf{x}'}\exp\left[-\frac{(\mathbf{x}' - \mathbf{x})^2}{2\sigma^2\mathbf{x}}\right]p(\mathbf{x})d\mathbf{x}$$
$$= \int \left[\frac{\mathbf{x} - \mathbf{x}'}{\sigma^2\mathbf{x}}\right]\frac{1}{\sqrt{2\pi\sigma^2\mathbf{x}}}\exp\left[-\frac{(\mathbf{x}' - \mathbf{x})^2}{2\sigma^2\mathbf{x}}\right]p(\mathbf{x})d\mathbf{x}. \tag{23}$$

---

[1]From Eq. (12), the approximation becomes accurate as $e_t^{\text{annea}} \to 0$, where $\tilde{\mathbf{x}}_t \to \mathbf{x}$.

Inserting the definition of $p(\mathbf{x'}|\mathbf{x})$ given in Eq. (20) to Eq. (23) herein results in

$$\nabla_{\mathbf{x'}} p(\mathbf{x'}) = \int \left[\frac{\mathbf{x} - \tilde{\mathbf{x}}}{\sigma^2 \mathbf{x}} \right] p(\mathbf{x'}|\mathbf{x}) p(\mathbf{x}) d\mathbf{x} \ . \quad (24)$$

Divide both sides of Eq. (24) herein by $p(\mathbf{x'})$. This leads to

$$\frac{\nabla_{\mathbf{x'}} p(\mathbf{x'})}{p(\mathbf{x'})} = \int \left[\frac{\mathbf{x} - \tilde{\mathbf{x}}}{\sigma^2 \mathbf{x}} \right] \frac{p(\mathbf{x'}|\mathbf{x}) p(\mathbf{x})}{p(\mathbf{x'})} d\mathbf{x} \ . \quad (25)$$

Recall Bayes rule,

$$p(\mathbf{x}|\mathbf{x'}) = \frac{p(\mathbf{x'}|\mathbf{x}) p(\mathbf{x})}{p(\mathbf{x'})} \ . \quad (26)$$

Inserting Eq. (26) in Eq. (25) herein results in

$$\frac{\nabla_{\mathbf{x'}} p(\mathbf{x'})}{p(\mathbf{x'})} = \int \left[\frac{\mathbf{x} - \tilde{\mathbf{x}}}{\sigma^2 \mathbf{x}} \right] p(\mathbf{x}|\mathbf{x'}) d\mathbf{x} \ . \quad (27)$$

The left-hand side of Eq. (27) herein satisfies

$$\frac{\nabla_{\mathbf{x'}} p(\mathbf{x'})}{p(\mathbf{x'})} = \nabla_{\mathbf{x'}} \log p(\tilde{\mathbf{x}}) \ . \quad (28)$$

The right-hand side of Eq. (27) herein can be given by

$$\int \left[\frac{\mathbf{x} - \mathbf{x'}}{\sigma^2 \mathbf{x}} \right] p(\mathbf{x}|\mathbf{x'}) d\mathbf{x} = \mathbb{E} \left[\frac{\mathbf{x} - \mathbf{x'}}{\sigma^2 \mathbf{x}} \Big| \mathbf{x'} \right] \ . \quad (29)$$

Overall, from Eqs. (27) to (29) herein,

$$\nabla_{\mathbf{x'}} \log p(\mathbf{x'}) = \mathbb{E} \left[\frac{\mathbf{x} - \mathbf{x'}}{\sigma^2 \mathbf{x}} \Big| \mathbf{x'} \right] \ . \quad (30)$$

The expectation in Eq. (30) herein cannot be calculated analytically. The reason is that the probability distribution of $\mathbf{x}$ is not accessible. Therefore, we approximate the expectation in Eq. (30) herein using a DNN. We generate a large dataset of synthetic pairs of clean and noisy images $(\mathbf{x}, \mathbf{x'})$ using Eq. (19) herein, with various standard deviations σ. Then, we train a deep neural network $h_\Theta^{\text{Poisson}}(\mathbf{x'}|\sigma)$ with learnable parameters Θ, such that

$$\hat{\Theta} = \arg \min_\Theta \mathbb{E} \left[\left\| \frac{\mathbf{x} - \mathbf{x'}}{\sigma^2 \mathbf{x}} - h_\Theta^{\text{Poisson}}(\mathbf{x'}, \sigma) \right\|_2^2 \right] \ . \quad (31)$$

Overall, using Eqs. (6), (18), (30) and (31) herein

$$\nabla_{\tilde{\mathbf{x}}_t} \log p(\tilde{\mathbf{x}}_t|\mathbf{y}) =$$
$$\frac{1}{2(\sigma_0^2 - \sigma_t^2)} \left(\frac{\mathbf{y}^2}{\tilde{\mathbf{x}}_t^2} - 1 \right) - \frac{1}{2\tilde{\mathbf{x}}_t} + h_{\hat{\Theta}}^{\text{Poisson}}(\tilde{\mathbf{x}}_t|\sigma_t) \ . \quad (32)$$

Algorithm 2: Langevin dynamics for complex-valued objects.

Input: $\{\sigma_t\}_{t=1}^T, \mathbf{y}, \mathbf{H}, \sigma_0$ and a small constant ϵ
Initialize $\tilde{\mathbf{o}}_0$, e.g. by Gaussian noise, or by $\mathbf{y}$
for $t \leftarrow 1$ *to* T **do**
$\quad \alpha_t \leftarrow \epsilon \frac{\sigma_t^2}{\sigma_T^2}$
$\quad$ Draw $\mathbf{n}_t^{\text{Langevin}} = N(0, \mathbf{I}) + j N(0, \mathbf{I})$
$\quad \boldsymbol{\Delta}_t \leftarrow \nabla_{\tilde{\mathbf{o}}_{t-1}} \log p(\mathbf{y}|\tilde{\mathbf{o}}_{t-1}) + h_{\hat{\Theta}}^{\text{complex}}(\tilde{\mathbf{o}}_{t-1}, \sigma_t)$
$\quad \tilde{\mathbf{o}}_t \leftarrow \tilde{\mathbf{o}}_{t-1} + \alpha_t \boldsymbol{\Delta}_t + \sqrt{2\alpha_t} \mathbf{n}_t^{\text{Langevin}}$
end
$\hat{\mathbf{o}} \leftarrow \tilde{\mathbf{o}}_T$

---

2. Recovering Complex-Valued Objects

This section details the complete derivations of annealed Langevin dynamics for complex valued, nonlinear imaging models, as described in Sec. 3.2 of the main manuscript. To be self-contained, this section repeats some equations that already exist in the main manuscript. Here, a noisy measurement satisfies

$$\mathbf{y} = \mathcal{N}(|\mathbf{Ho}|^2) \ , \quad (33)$$

where cross this paper, $|\cdot|^2$ operates element-wise. For analysis, we act as if $\mathbf{y}$ is related by Eq. (8) of the main manuscript, that is,

$$\mathbf{y} = |\mathbf{Ho}|^2 + \mathbf{n}^{\text{meas}} \ , \quad (34)$$

where

$$\mathbf{n}^{\text{meas}} \sim N(0, \sigma_0^2 |\mathbf{Ho}|^2) \ . \quad (35)$$

Let $\Re(\cdot)$, $\Im(\cdot)$ be real and imaginary arguments of a complex-valued number, respectively. Analogously to Eq. (5) herein, annealed Langevin dynamics for a complex-valued object is

$$\tilde{\mathbf{o}}_{t+1} = \tilde{\mathbf{o}}_t + \alpha_t \nabla_{\tilde{\mathbf{o}}_t} \log p(\tilde{\mathbf{o}}_t|\mathbf{y}) + \sqrt{2\alpha_t} \mathbf{n}_t^{\text{Langevin}} \ . \quad (36)$$

Using Bayes rule,

$$\nabla_{\tilde{\mathbf{o}}_t} \log p(\tilde{\mathbf{o}}_t|\mathbf{y}) = \nabla_{\tilde{\mathbf{o}}_t} [\log p(\mathbf{y}|\tilde{\mathbf{o}}_t) + \log p(\tilde{\mathbf{o}}_t)] \ . \quad (37)$$

Here, $\Re(\mathbf{n}_t^{\text{Langevin}}), \Im(\mathbf{n}_t^{\text{Langevin}}) \sim N(0, \mathbf{I})$, are distributed independently. In Section 2.1 herein we detail $\nabla_{\tilde{\mathbf{o}}_t} \log p(\mathbf{y}|\tilde{\mathbf{o}}_t)$, while $\nabla_{\tilde{\mathbf{o}}_t} \log p(\tilde{\mathbf{o}}_t)$ is described in Section 2.2 herein. Overall, the algorithm for Langevin dynamics for complex-valued objects is shown in Algorithm 2 herein. The small constant ϵ in Algorithm 2 herein is a predefined step size.

2.1. Likelihood Gradient: Complex-Valued Objects

Sec. 3.2 of the main manuscript presents the special case $\mathbf{H} = \mathbf{I}$ in Eq. (34) herein. Here, nevertheless we detail the

derivations for a general $\mathbf{H}$ matrix. In analogy to Eq. (11) herein, in Section 2 herein, we define a *complex-valued* dummy error term $\mathbf{e}_t^{\mathrm{annea}}$. Regardless of the exact value of $\mathbf{e}_t^{\mathrm{annea}}$, which is given in Section 2.1 herein, we set

$$\tilde{\mathbf{o}}_t = \mathbf{o} + \mathbf{e}_t^{\mathrm{annea}} \ . \tag{38}$$

By definition

$$|\mathbf{H}\mathbf{o}_t|^2 = [\Re(\mathbf{H}\mathbf{o}_t)]^2 + [\Im(\mathbf{H}\mathbf{o}_t)]^2 \ , \tag{39}$$

and

$$|\mathbf{H}\mathbf{e}_t^{\mathrm{annea}}|^2 = [\Re(\mathbf{H}\mathbf{e}_t^{\mathrm{annea}})]^2 + [\Im(\mathbf{H}\mathbf{e}_t^{\mathrm{annea}})]^2 \ . \tag{40}$$

Let $\odot$ denote element-wise product. Define

$$\mathbf{e}_t^{\mathrm{couple}} = 2\Re(\mathbf{H}\tilde{\mathbf{o}}_t) \odot \Re(\mathbf{H}\mathbf{e}_t^{\mathrm{annea}})$$
$$+ 2\Im(\mathbf{H}\tilde{\mathbf{o}}_t) \odot \Im(\mathbf{H}\mathbf{e}_t^{\mathrm{annea}}) \ . \tag{41}$$

Using Eqs. (38) to (41) herein, $|\mathbf{H}\mathbf{o}|^2$ satisfies

$$|\mathbf{H}\mathbf{o}|^2 = |\mathbf{H}(\tilde{\mathbf{o}}_t - \mathbf{e}_t^{\mathrm{annea}})|^2 = |\mathbf{H}\tilde{\mathbf{o}}_t - \mathbf{H}\mathbf{e}_t^{\mathrm{annea}}|^2$$
$$= [\Re(\mathbf{H}\tilde{\mathbf{o}}_t) - \Re(\mathbf{H}\mathbf{e}_t^{\mathrm{annea}})]^2 + [\Im(\mathbf{H}\tilde{\mathbf{o}}_t) - \Im(\mathbf{H}\mathbf{e}_t^{\mathrm{annea}})]^2$$
$$= |\mathbf{H}\tilde{\mathbf{o}}_t|^2 - \mathbf{e}_t^{\mathrm{couple}} + |\mathbf{H}\mathbf{e}_t^{\mathrm{annea}}|^2 \ , \tag{42}$$

Overall, from Eqs. (34) and (42) herein,

$$\mathbf{y} = |\mathbf{H}\tilde{\mathbf{o}}_t|^2 - \mathbf{e}_t^{\mathrm{couple}} + |\mathbf{H}\mathbf{e}_t^{\mathrm{annea}}|^2 + \mathbf{n}^{\mathrm{meas}} \ . \tag{43}$$

The non-linearity of Eq. (34) herein yields two terms in Eq. (43) herein, that pose analytic challenges: a coupling term of $\mathbf{e}_t^{\mathrm{couple}}$; and a squared annealing error term $|\mathbf{H}\mathbf{e}_t^{\mathrm{annea}}|^2$. We handle them by introducing a *relaxed* model

$$\mathbf{y} = |\mathbf{H}\tilde{\mathbf{o}}_t|^2 - \mathbf{e}_t^{\mathrm{couple}} + |\mathbf{H}\mathbf{e}_t^{\mathrm{annea}}|^2 + \tilde{\mathbf{n}}_t^{\mathrm{meas}} - \sigma_t^2 \mathbf{z}_t \ . \tag{44}$$

We detail Eq. (44) herein and specifically $\mathbf{e}_t^{\mathrm{couple}}, \mathbf{e}_t^{\mathrm{annea}}, \tilde{\mathbf{n}}_t^{\mathrm{meas}}$ and $\mathbf{z}_t$ in Section 2.1.1 herein.

2.1.1 Relaxation of the Nonlinear Problem

Let

$$\Re(\mathbf{e}_\tau^{\mathrm{part}}), \Im(\mathbf{e}_\tau^{\mathrm{part}}) \sim N\left[0, \frac{1}{4}(\sigma_t^2 - \sigma_{t+1}^2)\mathbf{I}\right] \tag{45}$$

be statistically independent Gaussian vectors. Recall that a sum of zero-mean Gaussian variables is a Gaussian, whose variance is the sum of the variances. Hence, in analogy to Eq. (10) herein, and using Eq. (45) herein

$$\sum_{\tau=t_1}^{t_2} \Re(\mathbf{e}_\tau^{\mathrm{part}}), \sum_{\tau=t_1}^{t_2} \Im(\mathbf{e}_\tau^{\mathrm{part}}) \sim$$
$$N\left[0, \frac{1}{4}\sum_{\tau=t_1}^{t_2}(\sigma_\tau^2 - \sigma_{\tau+1}^2)\mathbf{I}\right] = N\left[0, \frac{1}{4}(\sigma_{t_1}^2 - \sigma_{t_2}^2)\mathbf{I}\right] \ . \tag{46}$$

Note that using Eq. (45) herein

$$\sum_{\tau=0}^{T} \Re(\mathbf{e}_\tau^{\mathrm{part}}), \sum_{\tau=0}^{T} \Im(\mathbf{e}_\tau^{\mathrm{part}}) \sim N\left(0, \frac{1}{4}\sigma_0^2\mathbf{I}\right) \ . \tag{47}$$

Define the random vectors

$$\mathbf{n}_1^{\mathrm{part}} \equiv 2\Re(\mathbf{H}\mathbf{o}) \odot \sum_{\tau=0}^{T} \Re(\mathbf{e}_\tau^{\mathrm{part}})$$
$$\sim 2\Re(\mathbf{H}\mathbf{o}) \odot N\left(0, \frac{1}{4}\sigma_0^2\mathbf{I}\right) = N\left(0, \sigma_0^2 [\Re(\mathbf{H}\mathbf{o})]^2\right) \ , \tag{48}$$

and

$$\mathbf{n}_2^{\mathrm{part}} \equiv 2\Im(\mathbf{H}\mathbf{o}) \odot \sum_{\tau=0}^{T} \Im(\mathbf{e}_\tau^{\mathrm{part}})$$
$$\sim 2\Im(\mathbf{H}\mathbf{o}) \odot N\left(0, \frac{1}{4}\sigma_0^2\mathbf{I}\right) = N\left(0, \sigma_0^2 [\Im(\mathbf{H}\mathbf{o})]^2\right) \ . \tag{49}$$

Recall that a sum of zero-mean Gaussian variables is a Gaussian, whose variance is the sum of the variances. So, from Eqs. (48) and (49) herein, we obtain

$$\mathbf{n}_1^{\mathrm{part}} + \mathbf{n}_2^{\mathrm{part}} \sim N\left(0, \sigma_0^2 [\Re(\mathbf{H}\mathbf{o})]^2 + \sigma_0^2 [\Im(\mathbf{H}\mathbf{o})]^2\right) \ . \tag{50}$$

By definition

$$|\mathbf{H}\mathbf{o}|^2 = [\Re(\mathbf{H}\mathbf{o})]^2 + [\Im(\mathbf{H}\mathbf{o})]^2 \ . \tag{51}$$

From Eqs. (35) and (51) herein, we obtain

$$\mathbf{n}^{\mathrm{meas}} \sim N(0, \sigma_0^2 |\mathbf{H}\mathbf{o}|^2)$$
$$= N\left(0, \sigma_0^2 [\Re(\mathbf{H}\mathbf{o})]^2 + \sigma_0^2 \Im[(\mathbf{H}\mathbf{o})]^2\right) \ . \tag{52}$$

Using Eqs. (50) and (52) herein, we can design dummy variables $\mathbf{e}_t^{\mathrm{part}}$ such that,

$$\mathbf{n}^{\mathrm{meas}} = \mathbf{n}_1^{\mathrm{part}} + \mathbf{n}_2^{\mathrm{part}}$$
$$= 2\Re(\mathbf{H}\mathbf{o}) \odot \sum_{\tau=0}^{T} \Re(\mathbf{e}_\tau^{\mathrm{part}}) + 2\Im(\mathbf{H}\mathbf{o}) \odot \sum_{\tau=0}^{T} \Im(\mathbf{e}_\tau^{\mathrm{part}}) \ . \tag{53}$$

Notice that Eq. (53) herein now has a form similar to the coupling term $\mathbf{e}_t^{\mathrm{couple}}$ in Eq. (41) herein.

Recall that we seek to derive $\nabla_{\tilde{\mathbf{o}}_t} \log p(\mathbf{y}|\tilde{\mathbf{o}}_t)$. This requires expressing $\mathbf{y}$ using terms (besides $\tilde{\mathbf{o}}_t$) whose probability distributions are known and independent of $\tilde{\mathbf{o}}_t$. In analogy to Eq. (14) herein, we can hypothetically insert Eqs. (38) and (53) in Eq. (44). This leads to an expression of $\mathbf{y}$ that contains statistically dependent terms of $\tilde{\mathbf{x}}_t$ that do

not cancel each other, therefore, $\nabla_{\tilde{\mathbf{o}}_t} \log p(\mathbf{y}|\tilde{\mathbf{o}}_t)$ cannot be derived. Hence, to set an explicit analytical expression of $\nabla_{\tilde{\mathbf{o}}_t} \log p(\mathbf{y}|\tilde{\mathbf{o}}_t)$, we approximate the measurement noise as if the signal dependency is proportional to $\tilde{\mathbf{o}}_t$ rather then $\mathbf{o}$,

$$\tilde{\mathbf{n}}_t^{\text{meas}} = 2\Re(\mathbf{H}\tilde{\mathbf{o}}_t)\odot\sum_{\tau=0}^{T}\Re(\mathbf{e}_\tau^{\text{part}})+2\Im(\mathbf{H}\tilde{\mathbf{o}}_t)\odot\sum_{\tau=0}^{T}\Im(\mathbf{e}_\tau^{\text{part}}). \tag{54}$$

Overall, when $t \to T$ then, $\tilde{\mathbf{n}}_t^{\text{meas}} \to \mathbf{n}^{\text{meas}}$. We continue to generalize the annealing error in Eq. (11) to a *complex-valued* annealing error: analogously to [1], we design $\mathbf{e}_t^{\text{part}}$ such that

$$\mathbf{H}\mathbf{e}_t^{\text{annea}} = \sum_{\tau=t}^{T}\Re(\mathbf{e}_\tau^{\text{part}}) + j\sum_{\tau=t}^{T}\Im(\mathbf{e}_\tau^{\text{part}})\,, \tag{55}$$

where $j = \sqrt{-1}$. Inserting $\mathbf{H}\mathbf{e}_t^{\text{annea}}$ from Eq. (55) into Eq. (41) herein yields

$$\mathbf{e}_t^{\text{couple}} =$$
$$2\Re(\mathbf{H}\tilde{\mathbf{o}}_t) \odot \sum_{\tau=t}^{T}\Re(\mathbf{e}_\tau^{\text{part}}) + 2\Im(\mathbf{H}\tilde{\mathbf{o}}_t) \odot \sum_{\tau=t}^{T}\Im(\mathbf{e}_\tau^{\text{part}})\,. \tag{56}$$

Using Eqs. (54) and (56) herein, define

$$\mathbf{e}_t^{\text{remain}}(\tilde{\mathbf{o}}_t) \equiv \tilde{\mathbf{n}}_t^{\text{meas}} - \mathbf{e}_t^{\text{couple}}$$
$$= 2\Re(\mathbf{H}\tilde{\mathbf{o}}_t) \odot \sum_{\tau=0}^{t-1}\Re(\mathbf{e}_\tau^{\text{part}}) + 2\Im(\tilde{\mathbf{H}\mathbf{o}}_t) \odot \sum_{\tau=0}^{t-1}\Im(\mathbf{e}_\tau^{\text{part}})\,. \tag{57}$$

Until here in this section, we detailed $\mathbf{e}_t^{\text{couple}}, \mathbf{e}_t^{\text{annea}}, \tilde{\mathbf{n}}_t^{\text{meas}}$ of Eq. (44) herein. From this point, we detail $\mathbf{z}_t$ that appears there.

The vector $\mathbf{z}_t$ is designed to address the problem of squared annealing error term $|\mathbf{H}\mathbf{e}_t^{\text{annea}}|^2$. Inserting Eq. (55) into Eq. (40) herein, we obtain

$$|\mathbf{H}\mathbf{e}_t^{\text{annea}}|^2 = [\Re(\mathbf{H}\mathbf{e}_t^{\text{annea}})]^2 + [\Im(\mathbf{H}\mathbf{e}_t^{\text{annea}})]^2$$
$$= \left[\sum_{\tau=t}^{T}\Re(\mathbf{e}_\tau^{\text{part}})\right]^2 + \left[\sum_{\tau=t}^{T}\Im(\mathbf{e}_\tau^{\text{part}})\right]^2\,. \tag{58}$$

A chi-squared distribution $\chi^2(\kappa)$ is defined by a sum of κ random variables: each independent, having a square-normal distribution [2]. Fig. 1 herein visualizes $\chi^2(\kappa)$. So, $|\mathbf{H}\mathbf{e}_t^{\text{annea}}|^2$ in Eq. (58) herein is distributed $\chi^2(2)$, up to scale

$$f_{\chi^2(2)}\left(\frac{4}{\sigma_t^2}|\mathbf{H}\mathbf{e}_t^{\text{annea}}|^2\right) = \frac{1}{2}\exp\left(-\frac{4}{2\sigma_t^2}|\mathbf{H}\mathbf{e}_t^{\text{annea}}|^2\right), \tag{59}$$
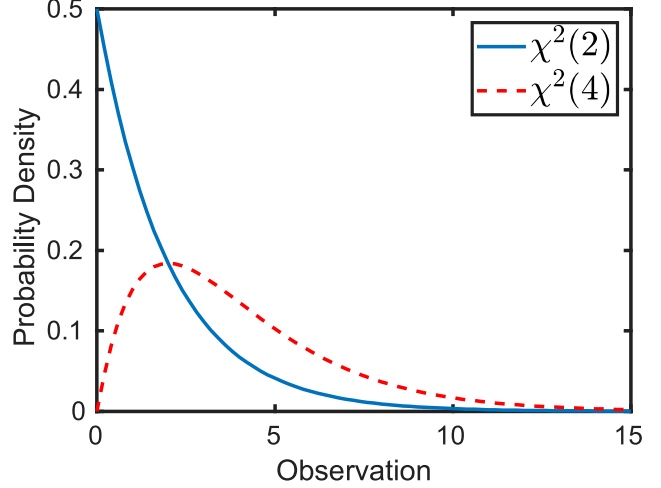

Figure 1: Visualization of $\chi^2(\kappa)$ for $\kappa = 2, 4$.

where $f_{\chi^2(2)}$ operates element-wise. Analogously to $|\mathbf{H}\mathbf{e}_t^{\text{annea}}|^2$, we propose two annealing error terms accordingly,[2]

$$\mathbf{z}_{t\to T} = \frac{\left[\Sigma_{\tau=t}^{T}\Re(\mathbf{e}_\tau^{\text{part}})\right]^2}{\sigma_t^2} + \frac{\left[\Sigma_{\tau=t}^{T}\Im(\mathbf{e}_\tau^{\text{part}})\right]^2}{\sigma_t^2}\,, \tag{60}$$

and

$$\mathbf{z}_{0\to(t-1)} = \frac{\left[\Sigma_{\tau=0}^{t-1}\Re(\mathbf{e}_\tau^{\text{part}})\right]^2}{\sigma_t^2} + \frac{\left[\Sigma_{\tau=0}^{t-1}\Im(\mathbf{e}_\tau^{\text{part}})\right]^2}{\sigma_t^2}\,. \tag{61}$$

So, $\mathbf{z}_{0\to(t-1)}, \mathbf{z}_{t\to T}$ in Eqs. (60) and (61) herein are distributed $\chi^2(2)$, up to scale. Then using Eqs. (60) and (61), we define

$$\mathbf{z}_t = \mathbf{z}_{t\to T} - \mathbf{z}_{0\to(t-1)}\,. \tag{62}$$

Using Eqs. (58) and (60) to (62) herein, we obtain

$$\mathbf{z}_t^{\text{remain}} \equiv |\mathbf{H}\mathbf{e}_t^{\text{annea}}|^2 - \sigma_t^2\mathbf{z}_t$$
$$= \left[\sum_{\tau=0}^{t-1}\Re(\mathbf{e}_\tau^{\text{part}})\right]^2 + \left[\sum_{\tau=0}^{t-1}\Im(\mathbf{e}_\tau^{\text{part}})\right]^2\,. \tag{63}$$

Inserting Eqs. (57) and (63) herein into the relaxed model in Eq. (44) herein, yields

$$\mathbf{y} = |\mathbf{H}\tilde{\mathbf{o}}_t|^2 + \mathbf{e}_t^{\text{remain}}(\tilde{\mathbf{o}}_t) + \mathbf{z}_t^{\text{remain}}\,. \tag{64}$$

Using Eq. (64) herein, $\nabla_{\tilde{\mathbf{o}}_t} \log p(\mathbf{y}|\tilde{\mathbf{o}}_t)$ can be derived analytically. To alleviate computations, we denote $\mathbf{s}$ as

$$\Re(\mathbf{s}) \equiv \sum_{\tau=0}^{t-1}\Re(\mathbf{e}_\tau^{\text{part}})\,, \ \Im(\mathbf{s}) \equiv \sum_{\tau=0}^{t-1}\Im(\mathbf{e}_\tau^{\text{part}})\,. \tag{65}$$
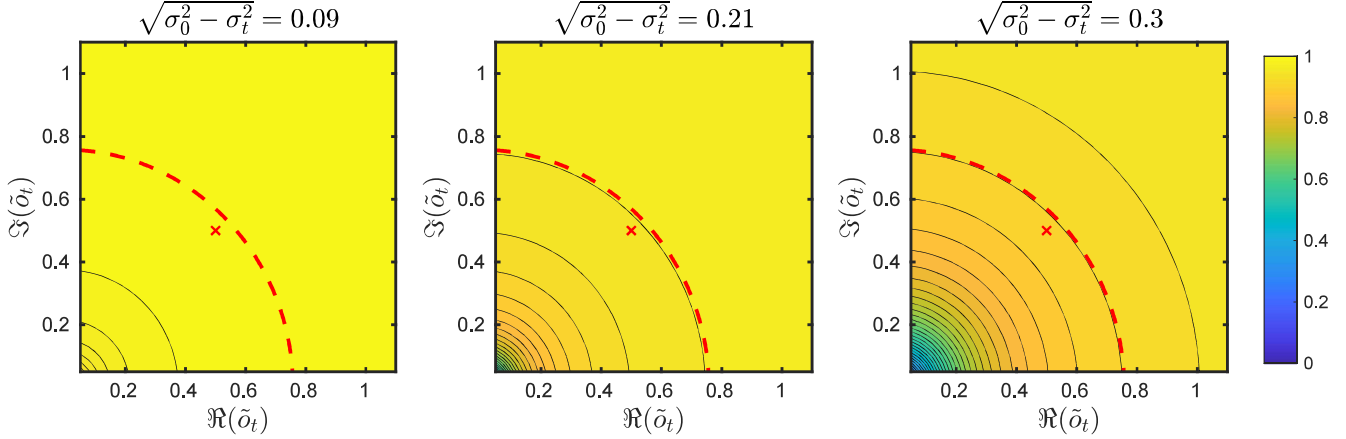
Figure 2: A visualization of $\mathcal{I}(\tilde{\mathbf{o}}_t, \mathbf{y})$ in Eq. (70) herein for different values of σ_t. Here $\mathbf{H} = 1$ and $o = 0.5 + 0.5j$, while y is obtained by Eq. (34) herein using $\sigma_0 = 0.3$. We present $\sqrt{y}$ by a red dashed line. The red cross marks o.
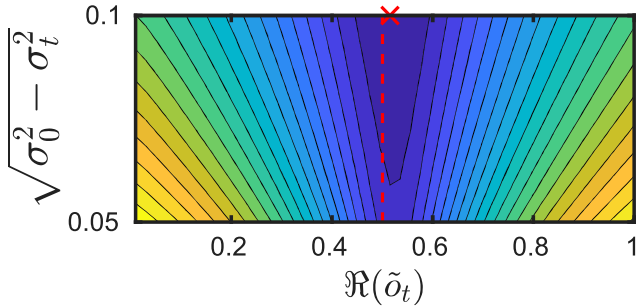

Figure 3: A visualization of the log-likelihood of Eq. (64) herein. Here $\mathbf{H} = 1$ and $o = 0.5 + 0.5j$. We present $\Re(o)$ by a red dashed line. Here, y is obtained by Eq. (34) herein, using $\sigma_0 = 0.1$. The red cross marks depicts the log-likelihood maximum.

Inserting Eq. (65) herein in Eq. (57) herein, yields

$$\mathbf{e}_t^{\mathrm{remain}}(\tilde{\mathbf{o}}_t) = 2\Re(\tilde{\mathbf{o}}_t) \odot \Re(\mathbf{s}) + 2\Im(\tilde{\mathbf{o}}_t) \odot \Im(\mathbf{s}) . \quad (66)$$

Inserting Eq. (65) herein in Eq. (63) herein, yields

$$\mathbf{z}_t^{\mathrm{remain}} = [\Re(\mathbf{s})]^2 + [\Im(\mathbf{s})]^2 . \quad (67)$$

Then, using $\mathbf{e}_t^{\mathrm{remain}}, \mathbf{z}_t^{\mathrm{remain}}$ from Eqs. (66) and (67) herein,

$$\mathbf{y} = |\mathbf{H}\tilde{\mathbf{o}}_t|^2 + 2\Re(\mathbf{H}\tilde{\mathbf{o}}_t) \odot \Re(\mathbf{s}) + 2\Im(\mathbf{H}\tilde{\mathbf{o}}_t) \odot \Im(\mathbf{s})$$
$$+ \left([\Re(\mathbf{s})]^2 + [\Im(\mathbf{s})]^2\right)$$
$$= [\Re(\mathbf{H}\tilde{\mathbf{o}}_t) + \Re(\mathbf{s})]^2 + [\Im(\mathbf{H}\tilde{\mathbf{o}}_t) + \Im(\mathbf{s})]^2 . \quad (68)$$

Notice that the probability distribution function of $(\Re(\mathbf{H}\tilde{\mathbf{o}}_t) + \Re(\mathbf{s}))^2 + (\Im(\mathbf{H}\tilde{\mathbf{o}}_t) + \Im(\mathbf{s}))^2$ in Eq. (68) is non-central chi-squared of the second kind, up to a scaling by

---

²Note that $\sigma_t^2 z_{t_1}$ in Eq. (44) herein can be considered as an additional annealing error in the measurement domain.

$(\sigma_0^2 - \sigma_t^2)$ [2]. Thus, using Eq. (68) herein, we can derive $p(\mathbf{y}|\tilde{\mathbf{o}}_t)$ analytically.

Let I_0, I_1 be the modified Bessel functions of the first kind [3], which operate element-wise. Let $\prod$ be an operator that outputs the element-wise product of a given vector. Then,

$$p(\mathbf{y}|\tilde{\mathbf{o}}_t) = \prod \left[\frac{1}{2(\sigma_0^2 - \sigma_t^2)} \exp\left(-\frac{\mathbf{y} + |\mathbf{H}\tilde{\mathbf{o}}_t|^2}{2(\sigma_0^2 - \sigma_t^2)}\right) \right.$$
$$\left. \odot I_0\left(\frac{\sqrt{|\mathbf{H}\tilde{\mathbf{o}}_t|^2 \odot \mathbf{y}}}{\sigma_0^2 - \sigma_t^2}\right) \right] . \quad (69)$$

We visualize $\log p(\mathbf{y}|\tilde{\mathbf{o}}_t)$ in Fig. 3 herein. Define,

$$\mathcal{I}(\tilde{\mathbf{o}}_t, \mathbf{y}) = I_1\left(\frac{|\mathbf{H}\tilde{\mathbf{o}}_t| \odot \sqrt{\mathbf{y}}}{\sigma_0^2 - \sigma_t^2}\right) \bigg/ I_0\left(\frac{|\mathbf{H}\tilde{\mathbf{o}}_t| \odot \sqrt{\mathbf{y}}}{\sigma_0^2 - \sigma_t^2}\right) . \quad (70)$$

Fig. 2 herein visualizes $\mathcal{I}(\tilde{\mathbf{o}}_t, \mathbf{y})$ in Eq. (70) herein. The complex-valued derivative of $\log p(\mathbf{y}|\tilde{\mathbf{o}}_t)$ is given by,

$$\nabla_{\tilde{\mathbf{o}}_t} \log p(\mathbf{y}|\tilde{\mathbf{o}}_t) =$$
$$\bar{\mathbf{H}}^\top \left(\frac{\mathbf{H}\tilde{\mathbf{o}}_t}{2(\sigma_0^2 - \sigma_t^2)} \odot \left[\mathcal{I}(\tilde{\mathbf{o}}_t, \mathbf{y}) \odot \frac{\sqrt{\mathbf{y}}}{|\mathbf{H}\tilde{\mathbf{o}}_t|} - 1 \right] \right) , \quad (71)$$

where $\bar{\mathbf{H}}^\top$ is the conjugate transpose of $\mathbf{H}$.

2.2. Gradient of the Prior: Complex-Valued Objects

We need to derive $\nabla_{\tilde{\mathbf{o}}_t} \log p(\tilde{\mathbf{o}}_t)$ in Eq. (37) herein for annealed Langevin dynamics of complex-valued objects. Analogously to Section 1.2 herein, before imaging takes place, we pre-train a DNN $h_\Theta^{\mathrm{Complex}}$ to approximate $\nabla_{\tilde{\mathbf{o}}_t} \log p(\tilde{\mathbf{o}}_t)$. In analogy to Eq. (38) herein, let $(\mathbf{o}, \mathbf{o}')$ be a synthetic pair of clean and noisy complex-valued objects, rendered using

$$\mathbf{o}' = \mathbf{o} + \mathbf{n}^{\mathrm{train}} . \quad (72)$$

Here $\Re(\mathbf{n}^{\text{train}}), \Im(\mathbf{n}^{\text{train}}) \sim N(0, \sigma^2 \mathbf{I})$, where σ is an arbitrary noise level. Using Eq. (72) herein,

$$p(\mathbf{o}'|\mathbf{o}) = p(\mathbf{o}' - \mathbf{o}|\mathbf{o}) = p(\mathbf{n}^{\text{train}}|\mathbf{o})$$
$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\mathbf{o}' - \mathbf{o})^2}{2\sigma^2}\right] . \quad (73)$$

By definition, the marginal distribution $p(\mathbf{o}')$ can be written as

$$p(\mathbf{o}') = \int p(\mathbf{o}'|\mathbf{o})p(\mathbf{o})d\mathbf{o} . \quad (74)$$

Inserting Eq. (73) into Eq. (74) herein yields

$$p(\mathbf{o}') = \int \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\mathbf{o}' - \mathbf{o})^2}{2\sigma^2}\right] p(\mathbf{o})d\mathbf{o} . \quad (75)$$

Then, the gradient of Eq. (75) herein satisfies

$$\nabla_{\mathbf{o}'}p(\mathbf{o}') = \int \frac{1}{\sqrt{2\pi\sigma^2}} \nabla_{\mathbf{o}'} \exp\left[-\frac{(\mathbf{o}' - \mathbf{o})^2}{2\sigma^2}\right] p(\mathbf{o})d\mathbf{o}$$
$$= \int \left[\frac{\mathbf{o} - \mathbf{o}'}{\sigma^2}\right] \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\mathbf{o}' - \mathbf{o})^2}{2\sigma^2}\right] p(\mathbf{o})d\mathbf{o} . \quad (76)$$

Inserting the definition of $p(\mathbf{o}'|\mathbf{o})$ given in Eq. (73) herein to Eq. (76) herein results in

$$\nabla_{\mathbf{o}'}p(\mathbf{o}') = \int \left[\frac{\mathbf{o} - \mathbf{o}'}{\sigma^2}\right] p(\mathbf{o}'|\mathbf{o})p(\mathbf{o})d\mathbf{o} . \quad (77)$$

Divide both sides of Eq. (77) herein by $p(\mathbf{o}')$. This leads to

$$\frac{\nabla_{\mathbf{o}'}p(\mathbf{o}')}{p(\mathbf{o}')} = \int \left[\frac{\mathbf{o} - \mathbf{o}'}{\sigma^2}\right] \frac{p(\mathbf{o}'|\mathbf{o})p(\mathbf{o})}{p(\mathbf{o}')}d\mathbf{o} . \quad (78)$$

Recall Bayes rule,

$$p(\mathbf{o}|\mathbf{o}') = \frac{p(\mathbf{o}'|\mathbf{o})p(\mathbf{o})}{p(\mathbf{o}')} . \quad (79)$$

Inserting Eq. (79) into Eq. (78) herein results in

$$\frac{\nabla_{\mathbf{o}'}p(\mathbf{o}')}{p(\mathbf{o}')} = \int \left[\frac{\mathbf{o} - \mathbf{o}'}{\sigma^2}\right] p(\mathbf{o}|\mathbf{o}')d\mathbf{o} . \quad (80)$$

The left-hand side of Eq. (80) herein satisfies

$$\frac{\nabla_{\mathbf{o}'}p(\mathbf{o}')}{p(\mathbf{o}')} = \nabla_{\mathbf{o}'} \log p(\mathbf{o}') . \quad (81)$$

The right-hand side of Eq. (80) herein can be given by,

$$\int \left[\frac{\mathbf{o} - \mathbf{o}'}{\sigma^2}\right] p(\mathbf{o}|\mathbf{o}')d\mathbf{o} = \mathbb{E}\left[\frac{\mathbf{o} - \mathbf{o}'}{\sigma^2}\bigg|\mathbf{o}'\right] . \quad (82)$$

Overall, from Eqs. (80) to (82) herein

$$\nabla_{\mathbf{o}'} \log p(\mathbf{o}') = \mathbb{E}\left[\frac{\mathbf{o} - \mathbf{o}'}{\sigma^2}\bigg|\mathbf{o}'\right] . \quad (83)$$

The expectation in Eq. (83) herein cannot be calculated analytically. The reason is that the probability distribution of $\mathbf{o}$ is not accessible. Therefore, we approximate the expectation in Eq. (83) herein using a DNN. We generate a large dataset of synthetic pairs of clean and noisy objects $(\mathbf{o}, \mathbf{o}')$ using Eq. (72) herein, with various standard deviations σ. Then, we train a deep neural network $h_\Theta^{\text{Complex}}(\mathbf{o}'|\sigma)$, such that

$$\hat{\Theta} = \arg\min_\Theta \mathbb{E}\left[\left\|\frac{\mathbf{o} - \mathbf{o}'}{\sigma^2} - h_\Theta^{\text{complex}}(\mathbf{o}'|\sigma)\right\|_2^2\right] . \quad (84)$$

Overall, using Eqs. (37), (71), (83) and (84) herein

$$\nabla_{\tilde{\mathbf{o}}_t} \log p(\tilde{\mathbf{o}}_t|\mathbf{y}) =$$
$$\frac{\tilde{\mathbf{o}}_t}{2(\sigma_0^2 - \sigma_t^2)} \odot \left[\mathcal{I}(\tilde{\mathbf{o}}_t, \mathbf{y}) \odot \frac{\sqrt{\mathbf{y}}}{|\tilde{\mathbf{o}}_t|} - 1\right] + h_{\hat{\Theta}}^{\text{complex}}(\tilde{\mathbf{x}}_t|\sigma_t) . \quad (85)$$

3. Fourier Ptychography: Additional Background

Here we describe a Fourier ptychography imaging system that was used in experiments [4]. Let $\mathbf{r}$ represent a 2D spatial coordinate. Let λ be the wavelength. The system has an array of M LEDs. Let $\boldsymbol{\delta}_m$ be the illumination angle of the mth LED. The model is that a LED creates illumination by a plane wave $u_m^{\text{LED}}(\mathbf{r})$. The plane wave angle $\boldsymbol{\delta}_m$ corresponds to a spatial frequency

$$\mathbf{k}_m = \sin(\boldsymbol{\delta}_m)/\lambda , \quad (86)$$

that is,

$$u_m^{\text{LED}}(\mathbf{r}) = \exp(j\mathbf{k}_m \cdot \mathbf{r}) . \quad (87)$$

Let $o(\mathbf{r})$ be the transmittance of an optically-thin object. The wave exiting the object is a multiplication of this transmittance and $u_m^{\text{LED}}(\mathbf{r})$ in Eq. (87) herein:

$$u_m^{\text{obj}}(\mathbf{r}) = o(\mathbf{r})\exp(j\mathbf{k}_m \cdot \mathbf{r}) . \quad (88)$$

Let $\mathcal{F}, \mathcal{F}^{-1}$ be the spatial Fourier transform and its inverse, respectively. Define

$$O(\mathbf{k}) = \mathcal{F}[o](\mathbf{k}) . \quad (89)$$

At the pupil plane (see Fig. 1 of the main manuscript), the light field corresponds to the spatial Fourier transform of $u_m^{\text{obj}}(\mathbf{r})$ in Eq. (88) herein,

$$U_m^{\text{obj}}(\mathbf{k}) = \mathcal{F}[u_m^{\text{obj}}](\mathbf{k}) = O(\mathbf{k} - \mathbf{k}_m) . \quad (90)$$

Let $P(\mathbf{k})$ be a low-pass filter, that is, a binary circle centered at the origin. Denote by k^{NA} the radius of $P(\mathbf{k})$, determined by the system numerical aperture (NA) and λ. Let $\mathcal{M}(\cdot)$ be a magnification operator of an imaging system.
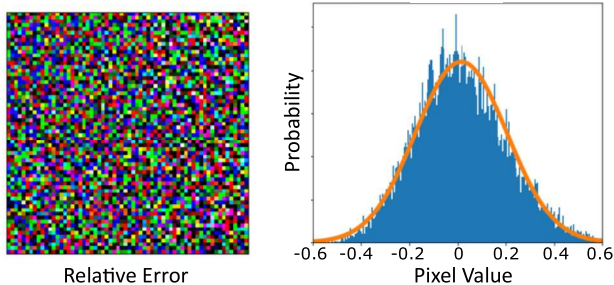
Figure 4: [Left] The relative error image $\hat{\mathbf{n}}^{\mathrm{rel}}$, following our recovery using data corrupted by $\sigma_0 = 0.2$. [Right] Normalized histogram of $\hat{\mathbf{n}}^{\mathrm{rel}}$. The orange line is a Gaussian curve, using the spatial mean and standard deviation $\mu_{\mathrm{rel}}, \sigma_{\mathrm{rel}}$, obtained by Eq. (92). Here, $\mu_{\mathrm{rel}} \approx 0$, and $\sigma_{\mathrm{rel}} = 0.19$.

The light intensity at the image plane resulting from a single LED is

$$y_m(\mathbf{r}) = \mathcal{N}\left(\mathcal{M}[|\mathcal{F}^{-1}\left[P(\mathbf{k})O(\mathbf{k} - \mathbf{k}_m)\right]|^2]\right)(\mathbf{r}) . \quad (91)$$

Hence, the angles of the LEDs and the support of $P(\mathbf{k})$ define the measured frequency content of the object. Fig. 12 herein visualizes the measured frequency content of the object for different numbers of LEDs.

4. Additional Results

In this section, we present additional results that were not included in the main manuscript, for space limitations. Furthermore, we detail and elaborate on the experiments presented in the main manuscript.

4.1. Denoising Poissonian Image Intensities

We follow the approach of [5] and assess if the estimated image holds $\hat{\mathbf{x}} \sim p(\mathbf{x})$. Denote a color channel by $c \in \{R, G, B\}$, a pixel index by $p \in [1, q^{\mathrm{meas}}]$, and a test image index by $i \in [1, N^{\mathrm{set}}]$. Define a relative error per channel c, pixel p and image i by $\hat{n}_{c,p}^{\mathrm{rel}}(i) \equiv [y_{c,p}(i) - \hat{x}_{c,p}(i)]/\sqrt{\hat{x}_{c,p}(i)}$. Using Eqs. (3) and (4), assessing if $\hat{\mathbf{x}} \sim p(\mathbf{x})$ can be done by verifying if a *relative error image* $\hat{\mathbf{n}}^{\mathrm{rel}} = [(\mathbf{y} - \hat{\mathbf{x}})/\sqrt{\hat{\mathbf{x}}}]$ is distributed as $N(0, \sigma_0^2 \mathbf{I})$. As shown in Fig. 4 herein, the distribution of $\hat{\mathbf{n}}^{\mathrm{rel}}$ seems indeed white and Gaussian. Let us assess the mean and variance of $\hat{n}_{c,p}^{\mathrm{rel}}$. Per i, the mean and the root mean squared relative
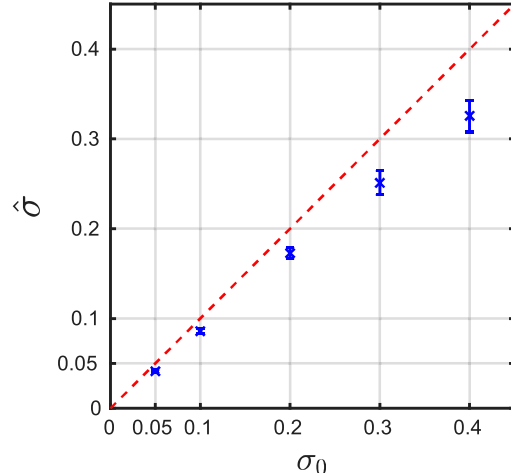

Figure 5: Poisson denoising. The blue crosses denote $\hat{\sigma}$ (Eq. 93), for different characterized noise levels σ_0. The dashed red line represents a perfect estimation.

errors are, respectively,

$$\mu_{\mathrm{rel}}(i) = \frac{1}{3q^{\mathrm{meas}}} \sum_{p=1}^{q^{\mathrm{meas}}} \sum_{c=1}^{3} \hat{n}_{c,p}^{\mathrm{rel}}(i) ,$$

$$\sigma_{\mathrm{rel}}(i) = \sqrt{\frac{1}{3q^{\mathrm{meas}}} \sum_{p=1}^{q^{\mathrm{meas}}} \sum_{c=1}^{3} [\hat{n}_{c,p}^{\mathrm{rel}}(i)]^2} . \quad (92)$$

We ran our method to denoise $N^{\mathrm{set}} = 1000$ images from the CelebA test set. For all denoised images, we computed $\hat{\mathbf{n}}^{\mathrm{rel}}(i)$ for $i \in [1, N^{\mathrm{set}}]$ and calculated $\mu_{\mathrm{rel}}(i), \sigma_{\mathrm{rel}}(i)$, using Eq. (92). We accurately obtained, $\forall i, \mu_{\mathrm{rel}}(i) \approx 0$. The distribution statistics of $\sigma_{\mathrm{rel}}(i)$ appear in Fig. 5 herein, including its empirical average over the test set, that is,

$$\hat{\sigma} = \frac{1}{N^{\mathrm{set}}} \sum_{i=1}^{N^{\mathrm{set}}} \sigma_{\mathrm{rel}}(i). \quad (93)$$

For $\sigma_0 \leq 0.2$, $\hat{\sigma}$ agrees with σ_0. This is supporting our quest for denoised images as if sampled from $p(\mathbf{x})$. For lower SNR scenarios, $\hat{\sigma}$ deviates.

Furthermore, we show additional examples of Poisson denoising. The experimental details are provided in the main manuscript. Figs. 7 to 9 herein compare our results, for different noise levels σ_0, with an MMSE solution. Our denoised images have higher quality than the MMSE solution, which is blurred. Figs. 10 and 11 herein compare our results, for different noise levels σ_0, with DPS [6] method.

4.2. Phase Retrieval

As described in Sec. 4.2 of the main manuscript, phase retrieval can be modeled using $\mathbf{H} = \mathbf{F}$, that is,

$$\mathbf{y} = \mathcal{N}(|\mathbf{Fo}|^2) . \quad (94)$$

Figure 6: Qualitative comparison between the ground-truth images (odd rows) in the dataset suggested by [7] and our recovery (even rows).

Here $\mathcal{N}$ contains photon noise and quantization, the latter using 8 bits per pixel. Moreover, here $\mathbf{o}$ is a real-valued image. Note that even so, the measurement $\mathbf{y}$ lacks phase information.

We compare our method (Algorithm 2 herein) for phase retrieval, vs. prior art. We follow the experiment suggested in [7]. First, we trained $h_\Theta^{\mathrm{Complex}}(\mathbf{o}'|\sigma)$ (Eq. 84 herein) with overlapping patches drawn from 400 images in the Berkeley Segmentation Dataset [8]. Then, we evaluate our method on a test set of 6 "natural" images, as presented in Fig. 6. We simulate for each image from the test set its noisy intensity measurements using Eq. (94) herein with noise level $\sigma_0 = 4/255$, as described in the main manuscript.

As reported in [7], the results were sensitive to initialization. So, as done in [7], we initialized our method using the result obtained by the Hybrid Input-Output (HIO) [9] algorithm. Fig. 13 presents a block diagram of the process. Let ψ be a random Gaussian vector. We initialized 50 possible candidate solutions $\{\mathbf{o}_l\}_{l=1}^{50}$ such that $\mathbf{F}\mathbf{o}_l = \sqrt{\mathbf{y}}\exp(j2\pi\psi_l)$. We ran HIO algorithm for 50 iterations for these candidates. The reconstructed candidate is the can-

didate with the lowest error

$$\left\| \mathbf{y} - |\mathbf{F}\mathbf{o}_l|_2^2 \right\|_2^2 . \tag{95}$$

Then, we initialize the HIO algorithm with the reconstructed candidate, and ran again the HIO algorithm for 1000 iterations, resulting in the HIO result. We initialize our method in Algorithm 2 herein with a noisy version of the HIO result. Following [7], this process was repeated three times and the reconstruction with the smallest error in Eq. (95) herein was used as the final estimate.

4.3. Fourier Ptychography: Real Experiment

In all experiments that are presented in the main manuscript and herein $\lambda = 512\mathrm{nm}$ and $\mathrm{NA} = 0.2$ [4]. For the configuration of $M = 89$ LEDs (see Fig. 12 herein), the maximal LED angle is set such that the effective numerical aperture of the system is $3.1 \times \mathrm{NA}$.

Fig. 12 herein visualizes the measuring coverage for different numbers of LEDs of the object's Fourier space in our experiments. We show Fourier ptychography results for real-world publicly available amplitude-only biological sample [10]. Using a Fourier ptychography imaging system, the amplitude of a sample of stained dog stomach cells was recovered in high resolution [4]. We randomly divided the recovered sample into non-spatially overlapping patches, each of size 256×256 pixels. Then, the patches were split randomly into training and test sets. We train our denoiser over the training set, according to Section 2 herein. For each patch $\mathbf{o}$ in the test set, we simulate M noisy measurements $\mathbf{y}$, each with a single LED. Figs. 14 and 15 herein show quantitative and qualitative results, respectively. Additionally, Fig. 16 herein shows qualitative recovery results of Fig. 9 from the main manuscript.

References

[1] Bahjat Kawar, Gregory Vaksman, and Michael Elad. SNIPS: Solving noisy inverse problems stochastically. *Advances in Neural Information Processing Systems (NIPS)*, 34:21757–21769, 2021. 5

[2] Henry Oliver Lancaster and Eugene Seneta. Chi-square distribution. *Encyclopedia of Biostatistics*, 2, 2005. 5, 6

[3] Frank Bowman. *Introduction to Bessel Functions.* Courier Corporation, 2012. 6

[4] Lei Tian, Xiao Li, Kannan Ramchandran, and Laura Waller. Multiplexed coded illumination for Fourier ptychography with an LED array microscope. *Biomedical Optics Express*, 5(7):2376–2389, 2014. 7, 9

[5] Bahjat Kawar, Gregory Vaksman, and Michael Elad. Stochastic image denoising by sampling from the posterior distribution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 1866–1875, 2021. 8

[6] Hyungjin Chung, Jeongsol Kim, Michael T Mccann, Marc L Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. *arXiv preprint arXiv:2209.14687*, 2022. 8

[7] Christopher Metzler, Phillip Schniter, Ashok Veeraraghavan, and Richard Baraniuk. Prdeep: Robust phase retrieval with a flexible deep network. In *International Conference on Machine Learning*, pages 3501–3510. PMLR, 2018. 9, 13

[8] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings Eighth IEEE International Conference on Computer Vision (ICCV)*, volume 2, pages 416–423. IEEE, 2001. 9

[9] James R Fienup. Phase retrieval algorithms: A comparison. *Applied Optics*, 21(15):2758–2769, 1982. 9, 13

[10] Lei Tian, Xiao Li, Kannan Ramchandran, and Laura Waller. *(Dataset) LED array 2D Fourier Ptychography. Available online.* http://gigapan.com/gigapans/170956. 9

Figure 7: CelebA Poisson denoising results for $\sigma_0 = 0.1$.

Figure 8: CelebA Poisson denoising results for $\sigma_0 = 0.2$.

Figure 9: CelebA Poisson denoising results for $\sigma_0 = 0.3$.

Figure 10: FFHQ Poisson denoising results for $\sigma_0 = 0.1$.

Figure 11: FFHQ Poisson denoising results for $\sigma_0 = 0.2$.
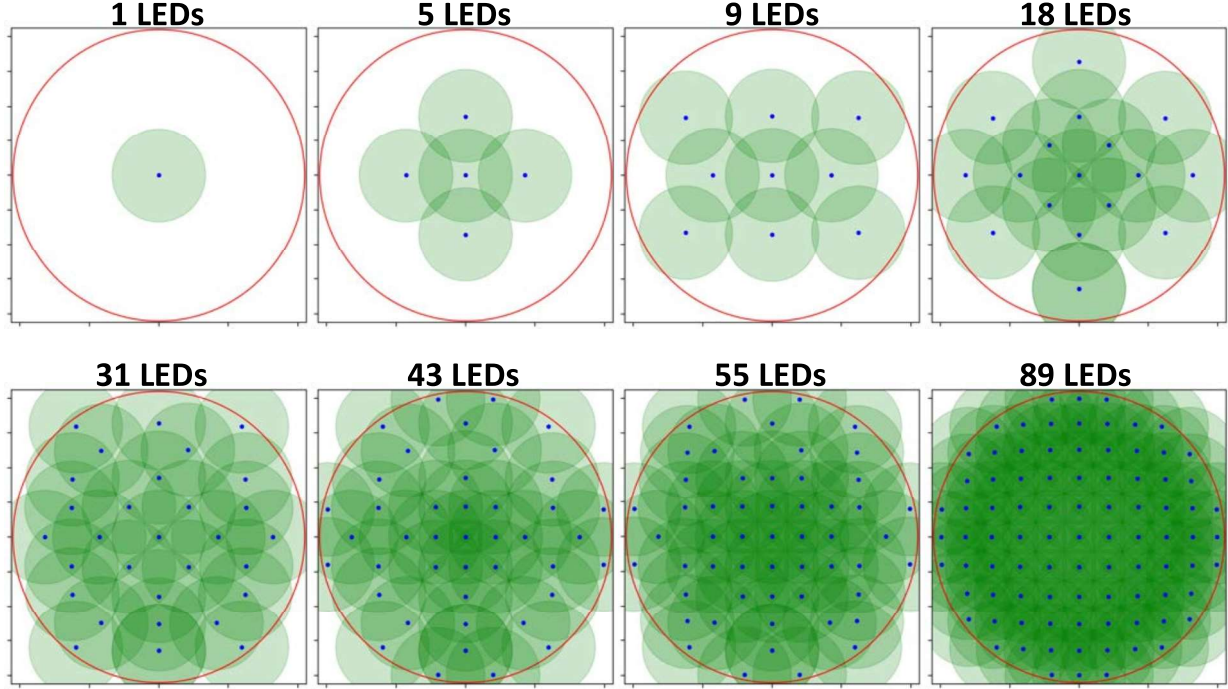
Figure 12: Visualization of coverage of the object's Fourier domain. The red circle is determined by the maximal LED angle of the system, having a frequency radius of k^{max}. The blue dots correspond to $\mathbf{k}_m$ of the LEDs, while the half-transparent green circles are the measured spectra domains around each $\mathbf{k}_m$, having frequency radius of k^{NA}.
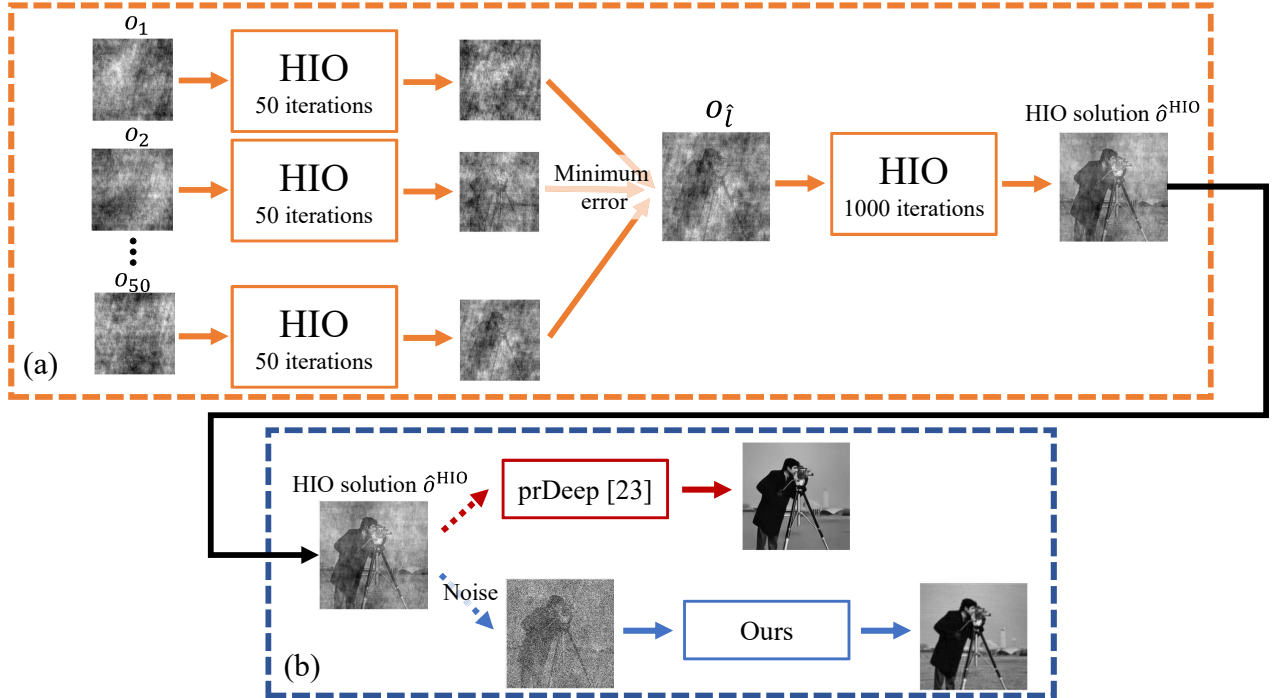

Figure 13: A block diagram of the phase retrieval experiment. (a) Initialization using the HIO [9] algorithm. (b) Final recovery using prDeep [7] and our method (Algorithm 2). The process of (a) and (b) is repeated three times, resulting in three solutions for each method $\{\hat{\mathbf{o}}_\ell^{\mathrm{ours}}\}_{\ell=1}^3$, $\{\hat{\mathbf{o}}_\ell^{\mathrm{prDeep}}\}_{\ell=1}^3$. Then, for each method, the reconstruction with the smallest error in Eq. (95) (out of the three) is used as its final estimation.
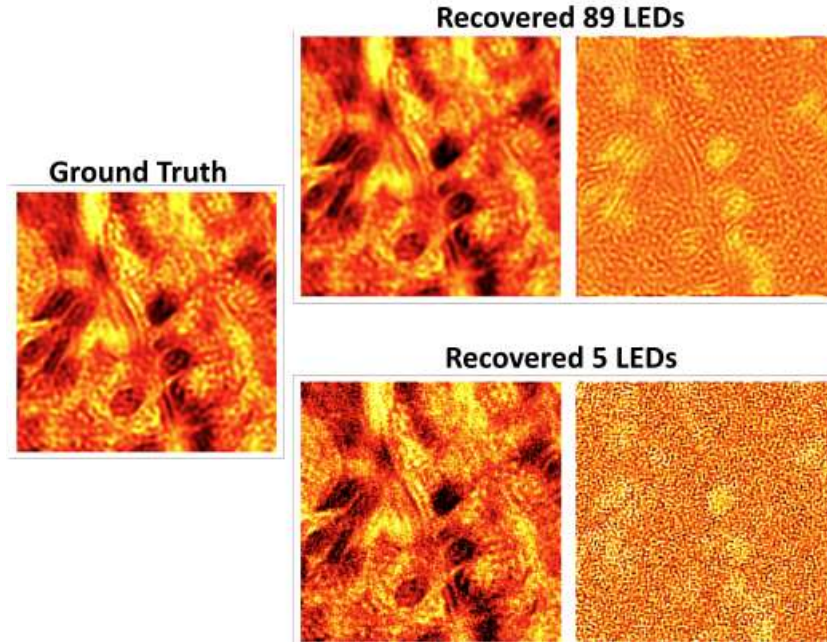
Figure 14: Qualitative results of amplitude objects. From left to right: Ground-truth image; Recovered images for $M = 5$ and $M = 89$ measurements; The difference between the true and estimated object. The dynamic range of the ground-truth and recovered images is $[0, 1]$. For better visualization, the dynamic range of both difference images is $[-0.04, 0.04]$.
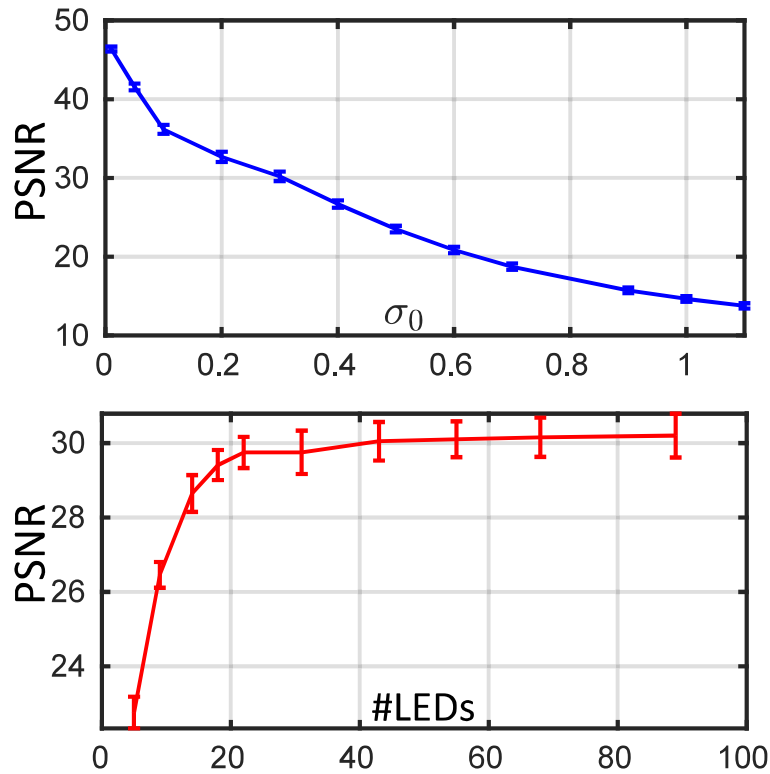

Figure 15: The average PSNR of 10 amplitude-only objects: [Top] For different Poisson noise levels σ_0 with $M = 89$. [Bottum] For a varying number of LEDs M with a fixed noise level $\sigma_0 = 0.3$. The bars stand for the PSNR standard deviation.
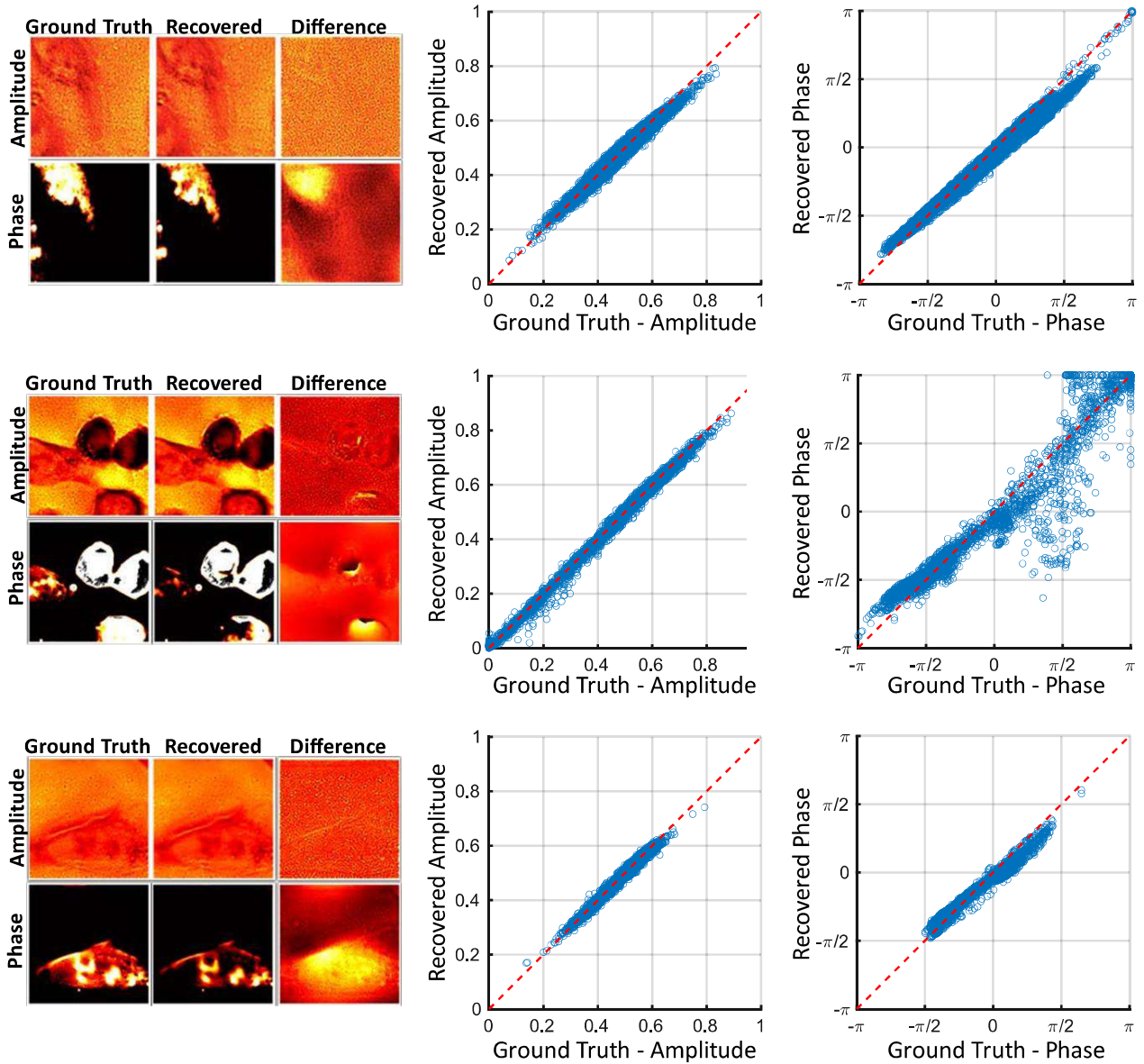
Figure 16: Fourier ptychography of complex-valued objects. Qualitative recovery results of four examples in our test set. Here, we used $M = 89$ LEDs, a noise level of $\sigma_0 = 0.1$, and quantization using 8 bits per pixel. The dynamic range of the ground-truth and recovered images of the amplitude and phase is $[0, 1]$ and $[-\pi, \pi]$, respectively. For the difference images, the dynamic range is tuned just for visualization.