Supplemental Document

Optics Letters

Maximum-likelihood estimation in ptychography in the presence of Poisson–Gaussian noise statistics: supplement

JACOB SEIFERT,[1,*] YIFENG SHAO,[1,2] RENS VAN DAM,[1] DORIAN BOUCHET,[3] TRISTAN VAN LEEUWEN,[4,5] AND ALLARD P. MOSK[1]

[1]*Nanophotonics, Debye Institute for Nanomaterials Science and Centre for Extreme Matter and Emergent Phenomena, Utrecht University, P.O. Box 80000, 3508 TA Utrecht, The Netherlands*
[2]*Imaging Physics Department, Applied Science Faculty, Delft University of Technology, The Netherlands*
[3]*Université Grenoble Alpes, CNRS, LIPhy, 38000 Grenoble, France*
[4]*Centrum Wiskunde & Informatica, Science Park 123, 1098 XG Amsterdam, The Netherlands*
[5]*Mathematical Institute, Utrecht University, Budapestlaan 6, 3584CD Utrecht, The Netherlands*
[*]*j.seifert@uu.nl*

Supplement DOI: https://doi.org/10.6084/m9.figshare.24421522

Parent Article DOI: https://doi.org/10.1364/OL.502344

Maximum-likelihood estimation in ptychography in the presence of Poisson-Gaussian noise statistics: supplemental document 1

This document (Supplement 1) provides supplementary material to *Maximum-likelihood estimation in ptychography in the presence of Poisson-Gaussian noise statistics*.

1. DERIVATION OF THE MAXIMUM-LIKELIHOOD ESTIMATION LOSS FUNCTIONS

In this supplementary chapter, we provide a derivation of the maximum-likelihood estimation (MLE) loss functions used in a ptychography framework based on automatic differentiation. MLE operates by selecting the set of parameters that maximize the likelihood function, thus ensuring the best fit to the observed data when the noise follows a known statistical probability distribution.

A. Poissonian noise statistics

Let us write the parameters of the ptychography model as a vector $\boldsymbol{\theta} = [\theta_1, \theta_2, ..., \theta_{\hat{N}}]$ with $\hat{N}$ denoting the total number of free parameters, presented in our case as complex-valued object pixels. Within a physics-based forward model of ptychography (as detailed in [1]), we can denote the expected intensity value at a certain pixel k as $I_k(\boldsymbol{\theta})$. In essence, $I_k(\boldsymbol{\theta})$ is the noise-free predicted intensity given a specific parameter vector $\boldsymbol{\theta}$. Considering a discrete random variable Y_k, which characterizes the intensity measurement on a camera sensor, we assume a Poisson distribution with an expectation value of $I_k(\boldsymbol{\theta})$. The probability mass function is thereby given by

$$p(Y_k|\boldsymbol{\theta}) = \frac{I_k(\boldsymbol{\theta})^{Y_k}}{Y_k!} \exp\left(-I_k(\boldsymbol{\theta})\right),\ Y_{k\in\{1,2,...,N\}}. \tag{S1}$$

Assuming that the N measurements are statistically independent, we can express the likelihood function as

$$\mathcal{L}(\boldsymbol{\theta}) = \prod_k^N p(Y_k|\boldsymbol{\theta}). \tag{S2}$$

For the sake of computational convenience, we employ the log-likelihood function $\ell(\boldsymbol{\theta})$, as the natural logarithm preserves order while transforming the product into the following sum:

$$\ell(\boldsymbol{\theta}) = \ln \mathcal{L}(\boldsymbol{\theta}) = \sum_k^N \left(Y_k \ln I_k(\boldsymbol{\theta}) - I_k(\boldsymbol{\theta}) - \ln Y_k!\right). \tag{S3}$$

As suggested in Chapter 4.1 of [2], we can now find the second-order Taylor expansion in terms of $\sqrt{I_k(\boldsymbol{\theta})}$ at the point $\sqrt{I_k(\boldsymbol{\theta})} = \sqrt{Y_k}$ as

$$Y_k \ln I_k(\boldsymbol{\theta}) - I_k(\boldsymbol{\theta}) \approx -Y_k + Y_k \ln Y_k - 2\left(\sqrt{Y_k} - \sqrt{I_k(\boldsymbol{\theta})}\right)^2. \tag{S4}$$

In practice, we aim to minimize the negative log-likelihood function. This transformation leads to equivalent outcomes and enables us to implement the optimization problem using a reconstruction algorithm based on gradient descent minimization and automatic differentiation libraries such as TensorFlow [3]. Hence, ignoring all constant additive terms and all multiplicative constants, the MLE loss function for Poissonian noise statistics can be written as

$$L_{\text{Poisson}}(\boldsymbol{\theta}) = -\ell(\boldsymbol{\theta}) = \sum_{k=1}^N \left(\sqrt{Y_k} - \sqrt{I_k(\boldsymbol{\theta})}\right)^2. \tag{S5}$$

B. Gaussian noise statistics

Minimizing the mean squared error is a common approach to optimization problems and is mathematically closely related to using an MLE loss function with the assumption of Gaussian noise statistics. Even though that is not the ideal assumption for a random variable W_k representing an intensity measurement, it becomes practicable for a large number of detected photons or for cases where W_k can be modeled as a sum of a large number of independent, identically distributed variables, regardless of their underlying distributions (central limit theorem). Considering the random variable W_k following a Gaussian distribution with the mean $I_k(\boldsymbol{\theta})$ and constant variance σ^2, we can express the probability density function as

$$p(W_k|\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(W_k - I_k(\boldsymbol{\theta}))^2}{2\sigma^2}\right). \tag{S6}$$

In analogy to the previous section, we can then express the log-likelihood as

$$\ell(\boldsymbol{\theta}) = \ln \mathcal{L}(\boldsymbol{\theta}) = \ln \prod_k^N p(W_k|\boldsymbol{\theta}) \tag{S7}$$

$$\ell(\boldsymbol{\theta}) = -\frac{N}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{k=1}^{N}\left(W_k - I_k(\boldsymbol{\theta})\right)^2. \tag{S8}$$

Now, by neglecting the constant additive term and multiplicative constants, it becomes evident that $\ell(\boldsymbol{\theta})$ can be maximized by the least squares method. We can write the MLE loss function for Gaussian noise statistics as

$$L_{\text{Gaussian}}(\boldsymbol{\theta}) = \sum_{k=1}^{N}\left(W_k - I_k(\boldsymbol{\theta})\right)^2. \tag{S9}$$

C. Mixed Poisson-Gaussian noise statistics

Continuing from the previous sections, let us consider a random variable X_k as the intensity measurement on a camera sensor with readout noise. We express X_k as the sum over two random variables $X_k = Y_k + Z_k$, where Y_k follows a Poisson distribution of expectation value $I_k(\boldsymbol{\theta})$ (see equation S1) and Z_k follows a centered Gaussian distribution of variance σ_k^2. The probability density function for Z_k is given by

$$p_{\text{g}}(Z_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left(-\frac{Z_k^2}{2\sigma_k^2}\right). \tag{S10}$$

The random variable Y_k can be approximated as a Gaussian distribution with mean $I_k(\boldsymbol{\theta})$ and variance $I_k(\boldsymbol{\theta})$:

$$p_{\text{p}}(Y_k|\boldsymbol{\theta}) \simeq \frac{1}{\sqrt{2\pi I_k(\boldsymbol{\theta})}} \exp\left[-\frac{(Y_k - I_k(\boldsymbol{\theta}))^2}{2I_k(\boldsymbol{\theta})}\right]. \tag{S11}$$

We can express the probability density function of X_k as

$$p(X_k|\boldsymbol{\theta}) = \int_{-\infty}^{+\infty} p_{\text{p}}(\tau|\boldsymbol{\theta}) p_{\text{g}}(X_k - \tau)\,\mathrm{d}\tau \tag{S12}$$

$$p(X_k|\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi(I_k(\boldsymbol{\theta}) + \sigma_k^2)}} \exp\left[-\frac{(X_k - I_k(\boldsymbol{\theta}))^2}{2(I_k(\boldsymbol{\theta}) + \sigma_k^2)}\right]. \tag{S13}$$

In analogy to the case without readout noise above, we can now define the MLE loss function as the negative log-likelihood function:

$$L_{\text{Mixed}}(\boldsymbol{\theta}) = -\ell(\boldsymbol{\theta}) = -\ln \mathcal{L}(\boldsymbol{\theta}) = -\ln \prod_k^N p(X_k|\boldsymbol{\theta}) \tag{S14}$$

$$L_{\text{Mixed}}(\boldsymbol{\theta}) = \sum_{k=1}^{N}\left(\ln[I_k(\boldsymbol{\theta}) + \sigma_k^2] + \frac{[X_k - I_k(\boldsymbol{\theta})]^2}{I_k(\boldsymbol{\theta}) + \sigma_k^2}\right). \tag{S15}$$

Here, we have neglected all constant additive terms and all multiplicative constants. To apply this loss function in an optimization framework, we need to obtain σ_k^2, the variance on the readout noise for each pixel, from dark measurements.

2. EXPERIMENTAL SETUP AND METHOD

The experimental setup for this ptychography study, also used for numerical simulations, is depicted in Fig 2 of the main document. A coherent laser beam (Cobolt Jive 100™) with wavelength $\lambda = 561\,\text{nm}$ is coupled into a single-mode fiber. A fiber collimator (60FC-L-0-M75-26, Schäfter+Kirchoff) expands the beam to around 25 mm in diameter, which then illuminates a 500 µm pinhole. Using a 2-lens system with a magnification of $M = 3$, the pinhole is imaged onto the object, resulting in the illumination field shown in Fig. S1, panel A. The two transfer lenses have diameters of 22.9 mm, with focal lengths of 5 mm and 15 mm, respectively. The object (µChart1951 Test Target, QingYing E&T LLC) is mounted on a motorized XY-stage with stepper motor actuators (ZFS25B, Thorlabs). The scanning trajectory is shown in panel B of Fig. S1. It comprises a total of 80 positions in a Fermat spiral pattern to optimize for overlap uniformness [4]. Using a traveling salesman algorithm, this trajectory is optimized to minimize total travel distance. The linear overlap between adjacent positions is approximately 60 % [5]. The diffraction patterns are recorded 37.7 mm behind the object using a CMOS camera (acA2440-35um, Basler) that features a binned pixel size of 6.9 µm and 1024x1024 total pixels.

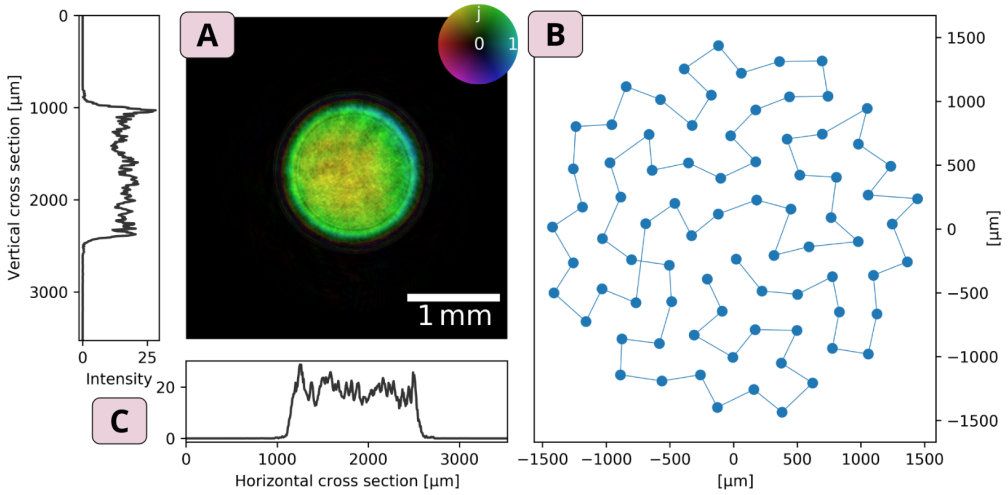

Fig. S1. (A) Visualization of the complex-valued illumination field that is used in the experiment. The image brightness represents the field amplitude, and the color represents the phase (see circular colorbar). (B) The object's scanning trajectory through the illumination beam in the ptychographic experiment. (C) Horizontal and vertical intensity profiles, centered on the illumination field.

To control the Signal-to-Noise Ratio (SNR) in the measurement, we vary the camera exposure time from 30 µs to 300 ms over 22 steps, covering four orders of magnitude. We derive the spatially varying readout noise variances σ_k^2 for each exposure time from 300 dark measurements, during which the laser beam is blocked. To mitigate Johnson-Nyquist noise fluctuations, we operate the camera sensor in a temperature controlled environment at 21 °C. These measurements also provide us with an average background image for each exposure time setting that we subtract from each diffraction pattern. To facilitate exact reproduction of the results presented in this study, the raw background and noise statistics data are included alongside the reconstruction framework in [6].

It is crucial to extract statistical information from potential negative intensity values resulting from Gaussian readout noise. Therefore, we require a black level offset to ensure that no pixel of the sensor ever reads the value of zero in dark measurements. With our Basler camera, we monitor the smallest pixel value for increasing black level settings and observe that an offset of 4 first ensures that all pixel values are larger than zero. To minimize the reduction in dynamic range, we choose this relatively small black level offset for the rest of this work.

For each scanning point, an independent measurement is obtained for each exposure time setting, and an additional high-SNR measurement is taken by averaging 100 images with the highest exposure time. This high-SNR measurement aids the calibration during the reconstruction phase. In Fig. S2 presents an expanded version of Fig. 3 from the main manuscript. In the

left column, the noise degradation of a single diffraction pattern is shown for all 22 exposure time settings. The central and right columns (B and C) provide a visual comparison of the reconstruction quality for each of these exposure time settings. Specifically, column B showcases reconstructions obtained by using the Poissonian log-likelihood loss function $L_{\text{Poisson}}(\boldsymbol{\theta})$ for optimization, while column C displays reconstructions achieved by employing the mixed Poisson-Gaussian log-likelihood loss function $L_{\text{Mixed}}(\boldsymbol{\theta})$. This comparative illustration provides a clear understanding of the impact of the chosen loss function on the quality of reconstruction across a large range of exposure times.

3. RECONSTRUCTION PROCEDURE

The reconstruction procedure begins with diffraction pattern preprocessing. An experimentally acquired mean dark image is subtracted to correct for background noise and account for hot or dead pixels. In cases where L_{Poisson} is used for optimization, negative values are set to zero due to the need for a real-valued loss function. Negative values cannot be incorporated into a noise model that solely assumes Poissonian statistics.

Typical CCD and CMOS cameras involve an analog-to-digital converter that converts the number of detected photons into analog-to-digital units (ADU). To rectify the assumption that the intensity measurement is Poisson distributed, we rescale the data by the inverse of the overall system gain. In the case of our CMOS camera, the inverse of the overall system gain is specified by the manufacturer to be 2.7 $\frac{\text{e}^-}{\text{ADU}}$.

Initially, a high-SNR reconstruction is conducted on the calibration dataset discussed in Section 2. This helps rectify experimental uncertainties such as the object-camera distance and scanning positions, as well as obtaining a high-quality reconstruction of the illumination field (see Fig. S1, panel A). Following this, reconstructions from the lower-SNR diffraction patterns are retrieved using the pre-calibrated illumination field. Each reconstruction is performed in sequence on a commercial GPU (Nvidia RTX A6000) with the same hyperparameter and regularization settings. Over 100 epochs, the learning rate for the ADAM optimizer [7] starts at $lr = 0.1$ and exponentially decays at a rate $\lambda = 0.03$, following the schedule $lr_{\text{n}+1} = lr_{\text{n}}\text{e}^{-\lambda}$.

Three regularization terms are added to the loss function, resulting in a final loss function in the form of $L = L_{Poisson/Mixed} + \sum_{i=1}^{3} L_{\text{Reg,i}}$.

1. An L1 norm on the amplitudes for the illumination field outside a circular support constraint $\mathcal{S}$ with a radius of 1.5 mm. This regularization term accelerates the convergence of the illumination calibration and is motivated by our experimental setup producing a circular illumination with an approximate radius of 0.75 mm. This is expressed as

$$L_{\text{Reg,1}} = \alpha \sum_{(x,y)\in\mathcal{S}} |P(x,y)|, \tag{S16}$$

   where $P(x,y)$ denotes the 2-dimensional illumination field (P for "probe"). The factor α regulates the strength of the regularization, typically chosen as $\alpha \approx 100$ in our calibration procedure.

2. A minor L1 norm on the amplitudes in the object given by

$$L_{\text{Reg,2}} = \beta \sum_{(x,y)}^{\hat{N}} |O(x,y)|, \tag{S17}$$

   where $O(x,y)$ denotes the 2-dimensional complex-valued object with a total number of pixels $\hat{N}$. This regularization drives towards finding a compact solution and mitigates high object amplitudes in the object's boundary areas that are insufficiently illuminated. We set $\beta = 0.0001$.

3. A minor L1 norm on the summed magnitudes of the object in frequency space is expressed as

$$L_{\text{Reg,3}} = \gamma \sum_{(x,y)}^{\hat{N}} |\hat{O}(u,v)|, \tag{S18}$$

   where $\hat{O}(u,v)$ denotes the Fourier transform $\hat{O}(u,v) = \mathcal{FT}\{O(x,y)\}$. We observe that this regularization term can stabilize the optimization using L_{Mixed}, which sometimes exhibits a poorer convergence behavior than optimizing L_{Poisson} or helps prevent numerical divergence phenomena for less-than-optimal learning rates. We set $\gamma = 0.001$.

Note that choosing the regularization prefactors α, β, γ is arbitrary and depends on non-physical parameters such as the sampling. Therefore, we adopt a heuristical approach to set them small enough to ensure that the data fidelity term strongly dominates the reconstruction process. By doing so, we preserve the valuable comparative basis between the two maximum likelihood estimation (MLE) loss functions while subtly enhancing the reconstructions:

$$\frac{L_{\text{Poisson/Mixed}}}{\sum_{i=1}^{3} L_{\text{Reg,i}}} \geq 100. \tag{S19}$$

REFERENCES

1. J. Seifert, D. Bouchet, L. Loetgering, and A. P. Mosk, "Efficient and flexible approach to ptychography using an optimization framework based on automatic differentiation," OSA Continuum **4**, 121–128 (2021).
2. A. P. Konijnenberg, "Computational methods for phase retrieval," Ph.D. thesis, TU Delft (2019).
3. M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard *et al.*, "Tensorflow: a system for large-scale machine learning." in *Osdi*, vol. 16 (2016), pp. 265–283.
4. X. Huang, H. Yan, R. Harder, Y. Hwu, I. K. Robinson, and Y. S. Chu, "Optimization of overlap uniformness for ptychography," Opt. Express **22**, 12634–12644 (2014).
5. O. Bunk, M. Dierolf, S. Kynde, I. Johnson, O. Marti, and F. Pfeiffer, "Influence of the overlap parameter on the convergence of the ptychographical iterative engine," Ultramicroscopy **108**, 481–487 (2008).
6. J. Seifert, Y. Shao, and A. P. Mosk, "Supplemental code and raw data for 'maximum-likelihood estimation in ptychography in the presence of poisson-gaussian noise statistics'," Data publication platform Utrecht Univ. (2023). https://doi.org/10.24416/UU01-4SVE6U.
7. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980 (2014).

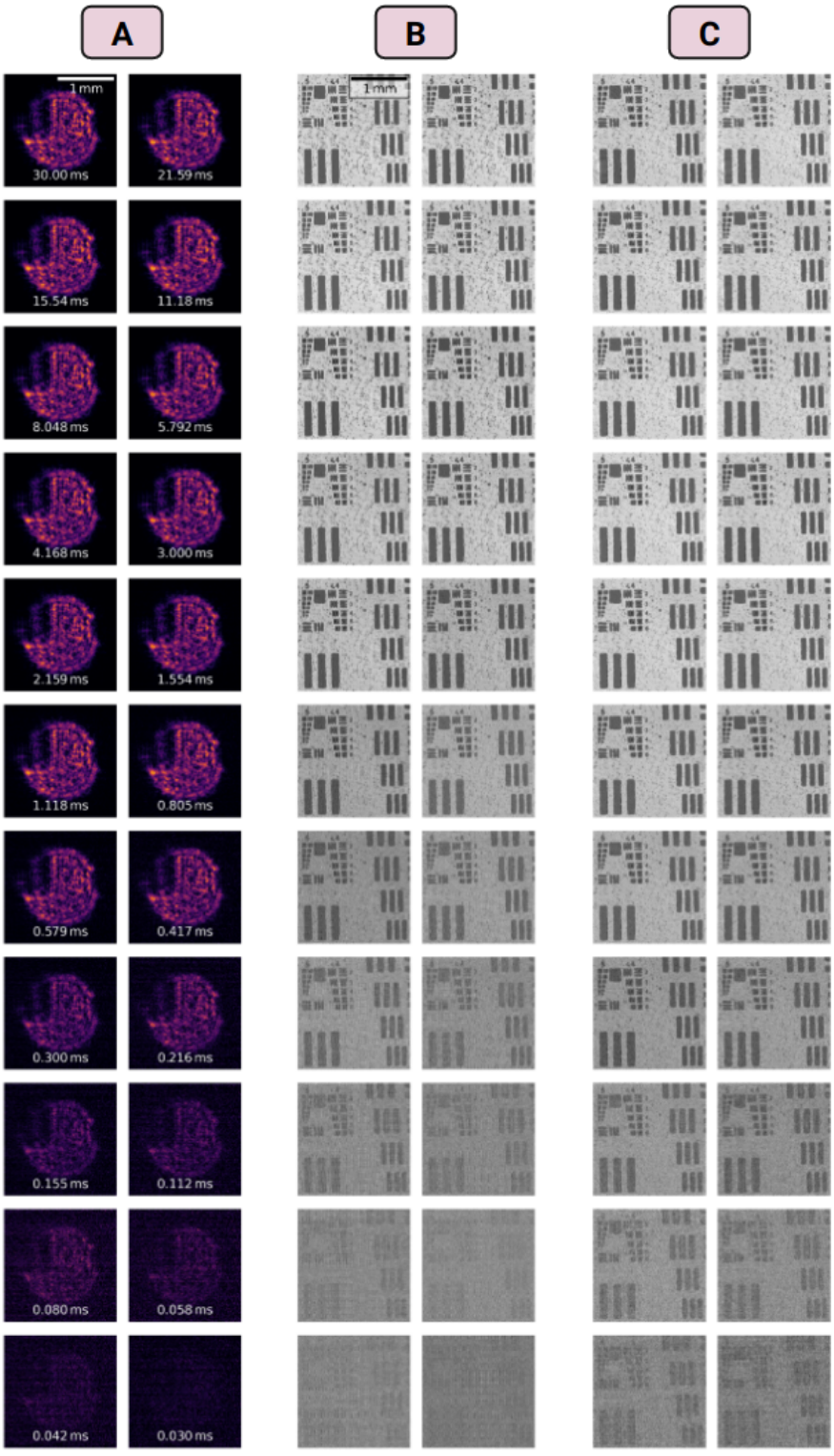

Fig. S2. Column (A): Visualization of one diffraction pattern from the full data set at varying camera exposure time settings. The exposure times range from 30 ms (upper left) to 30 µs (lower right). Columns (B and C): Amplitude images reconstructed from the ptychographic data sets with the respective exposure times shown in column A in the same order. Column B is reconstructed with a loss function assuming solely Poissonian noise statistics. Column C is reconstructed with the mixed-statistics loss function assuming Poisson-Gaussian noise.