2021 IEEE International Conference on Image Processing (ICIP) | 978-1-6654-4115-5/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICIP42928.2021.9506139

POISSON PHASE RETRIEVAL WITH WIRTINGER FLOW

Zongyu Li[†] *Kenneth Lange*[⋆] *Jeffrey A. Fessler*[†]

[†] University of Michigan – Ann Arbor
[⋆] University of California, Los Angeles

ABSTRACT

This paper discusses algorithms for phase retrieval where the measurements follow independent Poisson distributions. We developed an optimization problem based on maximum likelihood estimation (MLE) for the Poisson model and applied Wirtinger flow algorithm to solve it. Simulation results with a random Gaussian sensing matrix and Poisson measurement noise demonstrated that the Wirtinger flow algorithm based on the Poisson model produced higher quality reconstructions than when algorithms derived from Gaussian noise models (Wirtinger flow, Gerchberg Saxton) are applied to such data, with significantly improved computational efficiency.

***Index Terms*—** Poisson phase retrieval, Non-convex optimization.

1. INTRODUCTION

Phase retrieval is an inverse problem that has many applications in engineering and applied physics [1, 2], including radar [3], X-ray crystallography [4], astronomical imaging [5] and speech processing [6], where the goal is to recover the signal from only the magnitude of linear measurements, such as the magnitude of its Fourier transform [7].

In most prior works, the measurement vector $\boldsymbol{y} \in \mathbb{R}^M$ was assumed to have statistically independent elements with Gaussian distributions:

$$y_i \sim \mathcal{N}(|\boldsymbol{a}_i'\boldsymbol{x}|^2 + b_i, \sigma^2),$$

where $\boldsymbol{a}_i'$ denotes the ith row of the system matrix $\boldsymbol{A} \in \mathbb{C}^{M\times N}$, $\boldsymbol{x} \in \mathbb{C}^N$ denotes the true unknown signal, and b_i denotes a known mean background signal (often simply zero) for the ith measurement, where $i = 1, \ldots, M$. For the Gaussian noise model, the maximum-likelihood estimate of $\boldsymbol{x}$ corresponds to the following optimization problem:

$$\hat{\boldsymbol{x}} = \operatorname*{argmin}_{\boldsymbol{x}\in\mathbb{C}^N} f(\boldsymbol{x}), \quad f(\boldsymbol{x}) \triangleq \sum_i \left| y_i - b_i - \left|\boldsymbol{a}_i'\boldsymbol{x}\right|^2 \right|^2. \tag{1}$$

Many algorithms based on (1) have been proposed for phase retrieval, such as Wirtinger Flow [8], Gerchberg Saxton [9] and Majorizing Minimization [7]. However, in some low-photon count applications [10–14], a Poisson noise model is more appropriate:

$$y_i \sim \text{Poisson}(|\boldsymbol{a}_i'\boldsymbol{x}|^2 + b_i), \tag{2}$$

where here $b_i \geq 0$ denotes known mean background counts for the ith measurement, e.g., as arising from dark current [15]. Algorithms derived from Gaussian ML models are theoretically suboptimal in this case. Instead, the following Poisson ML model is more natural:

$$\begin{aligned} \hat{\boldsymbol{x}} =& \operatorname*{argmin}_{\boldsymbol{x}\in\mathbb{C}^N} f(\boldsymbol{x}), \quad f(\boldsymbol{x}) \triangleq \sum_i \psi(\boldsymbol{a}_i'\boldsymbol{x}; y_i, b_i), \\ &\psi(v; y, b) \triangleq (|v|^2 + b) - y\log(|v|^2 + b). \end{aligned} \tag{3}$$

Similar problems for the case $b_i = 0$ have been considered previously [16–21]. Many optical sensors also have Gaussian readout noise [22, 23]; the log likelihood for a Poisson + Gaussian distribution is complicated, so a common approximation is to use a shifted Poisson model that also leads to the cost function in (3). An alternative to the shifted Poisson model could be to work with an unbiased inverse transformation of a generalized Anscombe transform approximation [24]. This paper will focus on algorithms for the pure Poisson noise model (2); algorithms for a Poisson plus Gaussian noise model can be an interesting direction for future work.

However, one can verify that the function $h(|v|; y, b) = \psi(v; y, b)$ where $h(r; y, b) \triangleq (r^2 + b) - y\log(r^2 + b)$ is non-convex in $r \in \mathbb{R}$ when $0 < b < y$. That property, combined with the modulus within the logarithm, makes (3) a challenging optimization problem. One potential advantage of assuming $b_i > 0$ is the descent direction (4) of $\psi(v)$ is well defined, so that algorithms based on the gradient descent, *i.e.*, Wirtinger flow, are well-suited to "solve" (i.e., descend) this

Zongyu Li and Jeffrey A. Fessler are with Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109-2122 (e-mail: zonyul@umich.edu, fessler@umich.edu).

Kenneth Lange is with Departments of Computational Medicine, Human Genetics, and Statistics, University of California, Los Angeles, CA 90095 (e-mail: klange@ucla.edu).

Research supported in part by USPHS grants GM53275 and HG006139, and by NSF Grant IIS 1838179. A journal-length version of this paper with more algorithms, including regularized versions, can be found at `https://arxiv.org/abs/2104.00861`.

non-convex optimization problem. One can verify that an ascent direction for ψ for $v \in \mathbb{C}$ is

$$\dot{\psi}(v; y, b) = 2v\left(1 - \frac{y}{|v|^2 + b}\right). \tag{4}$$

2. WIRTINGER FLOW METHOD

Reference [8] proposed a Wirtinger flow algorithm for the Gaussian model (3) based on the steepest descent method with a heuristic step size. We adopted the steepest descent idea for the Poisson model (1), using a modified gradient term derived from (4). In addition, instead of empirically choosing the scheduling term μ, we did a line search as in [7] to seek the best value of μ. Algorithm 1 shows our modified Wirtinger flow (WF-Poisson) method, where the dot subscript denotes element-wise function application, as in Julia. For comparison, we also implemented the Gerchberg Saxton and Wirtinger flow methods that are both derived from Gaussian noise models (GS and WF-Gaussian). For WF-Gaussian, we simply modified the gradient term ∇f to be $4\boldsymbol{A}'\mathrm{diag}(|\boldsymbol{A}\boldsymbol{x}^{(k)}|^2 - \max(\boldsymbol{y} - \boldsymbol{b}, 0))\boldsymbol{A}\boldsymbol{x}^{(k)}$. For GS, we implemented the pseudo code shown in [7].

Algorithm 1: Wirtinger flow under Poisson noise

Input: $\boldsymbol{A}, \boldsymbol{y}, \boldsymbol{b}$ and n (number of iterations)
Initialize: $\boldsymbol{x}^{(0)} \leftarrow$ random Gaussian vector
for $k = 0, ..., n-1$ **do**
  $\nabla f^{(k)} = \boldsymbol{A}'\nabla\psi_.(\boldsymbol{A}\boldsymbol{x}^{(k)}; \boldsymbol{y}, \boldsymbol{b})$
  $\mu = 1$
  $\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \mu\nabla f^{(k)}$
  while $f(\boldsymbol{x}^{(k+1)}) > f(\boldsymbol{x}^{(k)}) - 0.01\,\mu\|\nabla f^{(k)}\|_2^2$ **do**
    $\mu \leftarrow \mu/2$
    $\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \mu\nabla f^{(k)}$
  end
end
Output: $\boldsymbol{x}^{(n)}$

By design, the Wirtinger flow method is guaranteed to decrease the cost function every iteration (strictly speaking it is non-increasing). Because the cost function is nonconvex, it is hard to say anything about convergence to a global minimizer. We find empirically (results not shown) that when the number of measurements M is close to N than the result of the WF method can depend significantly on the initial estimate $\boldsymbol{x}^{(0)}$, but if M is much larger than N then the NRMSE of the output of the WF method is insensitive to the initializer.

3. EXPERIMENTS

3.1. Experiment Setup

We adopted two finite length ($N = 100$) signals and one piece-wise uniform image (64×64) as the true signal/image in our experiments. The first finite length signal is a real, piece-wise constant signal ($\boldsymbol{x}_{\text{true-A}}$); the second is a complex signal ($\boldsymbol{x}_{\text{true-B}}$) whose real part is the same as $\boldsymbol{x}_{\text{true-A}}$ with piece-wise constant imaginary part. Fig. 1 shows $\boldsymbol{x}_{\text{true-A}}$, the imaginary part of $\boldsymbol{x}_{\text{true-B}}$ and the true image $\boldsymbol{x}_{\text{true-C}}$. To validate these algorithms in different experimental settings, we chose the number of measurements (M) from a set of numbers that range from 1e3 to 6e3 with an interval of 1e3. The system matrix $\boldsymbol{A}$ was a zero-mean complex Gaussian matrix, scaled by different constants such that the average of $|\boldsymbol{a}_i'\boldsymbol{x}_{\text{true}}|^2$ reached 2,3,4 and 5, respectively. The mean background counts ($\boldsymbol{b}$) were set to 1 in all cases. Elements in the measurement vector $\boldsymbol{y}$ were simulated to have independent Poisson distributions per (2). For $\boldsymbol{x}_{\text{true-A}}$ and $\boldsymbol{x}_{\text{true-B}}$, we initialized $\boldsymbol{x}$ as a Gaussian random vector and ran all algorithms for 250 iterations; for $\boldsymbol{x}_{\text{true-C}}$, the initial estimate $\boldsymbol{x}^{(0)}$ was a random vector following uniform distribution ranging from 0 to 1 (exploiting the nonnegativity of $\boldsymbol{x}$) and we ran all algorithms for 100 iterations to save computing time. All simulations ran on Mac OS with Intel Core i9@2.3 GHz CPU and 16 GB memory.

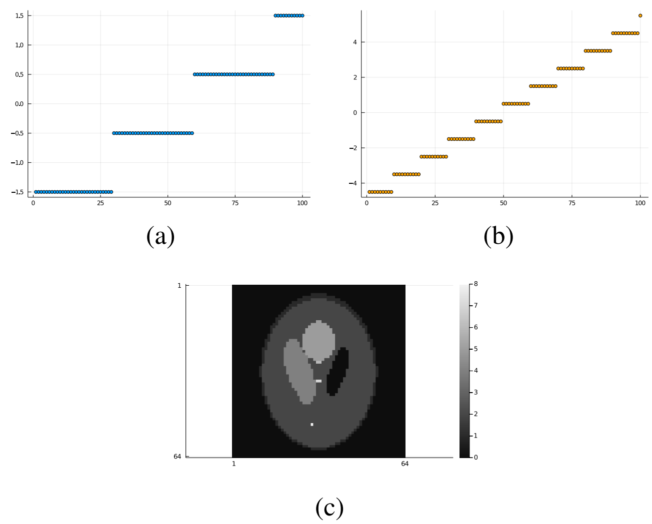


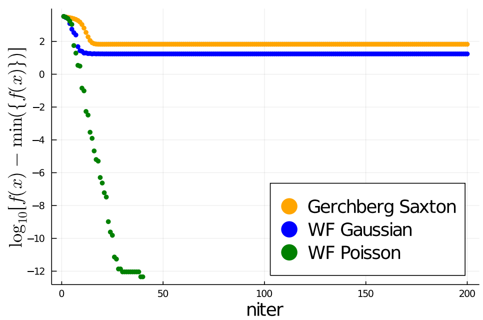
Fig. 1. True signals used in simulations. Subfigure (a) denotes a real, piece-wise constant signal $\boldsymbol{x}_{\text{true-A}}$; subfigure (b) denotes the imaginary part of the complex signal $\boldsymbol{x}_{\text{true-B}}$; subfigure (c) refers to the real, piece-wise uniform image $\boldsymbol{x}_{\text{true-C}}$.

3.2. Convergence Analysis

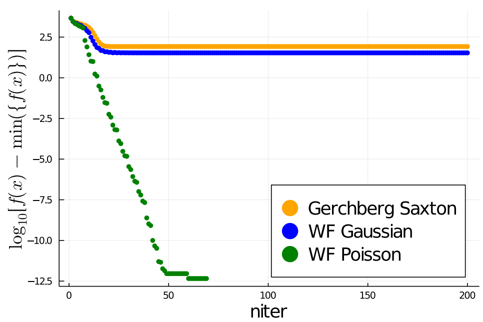
Fig. 2 illustrates the convergence rate of cost function value (3) versus number of iterations. The $\min\{f(\boldsymbol{x})\}$ in the y-axis label denotes the minimum of cost function values among all algorithms over 250 iterations. As expected, WF-Poisson achieved much lower cost function values at convergence compared to GS and WF-Gaussian. Fig. 3 shows the NRMSE versus wall-time for the three algorithms for a single noise realization. Compared to GS and WF-Gaussian, WF-Poisson decreased NRMSE much more rapidly.

To handle the phase ambiguity (all the algorithms can recover the signal only to within a constant phase shift due to the loss of global phase information), we used the following NRMSE metric as derived in [7]:

$$\text{NRMSE} = \frac{\|\hat{\boldsymbol{x}} - \boldsymbol{x}_{\text{true}} \cdot e^{j\phi}\|_2}{\|\boldsymbol{x}_{\text{true}} \cdot e^{j\phi}\|_2}, \quad e^{j\phi} = \text{sign}(\boldsymbol{x}'_{\text{true}}\hat{\boldsymbol{x}}). \quad (5)$$



(a) $\boldsymbol{x}_{\text{true-A}}$



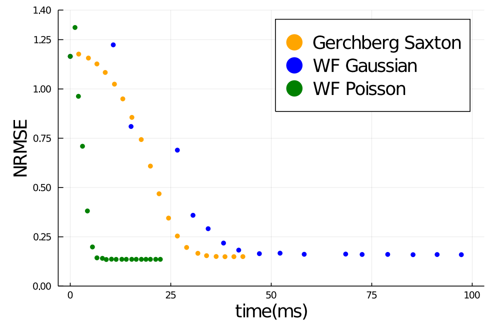
(b) $\boldsymbol{x}_{\text{true-B}}$

Fig. 2. Cost function value versus number of iterations. The average of $|\boldsymbol{a}_i'\boldsymbol{x}|^2$ is set to 2 and $M = 3000$.

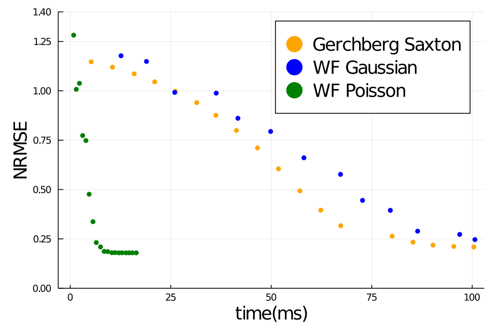
3.3. Accuracy comparison

As a more systematic investigation of phase retrieval accuracy, Fig. 4 shows the NRMSE averaged across 3 realizations of the sensing matrix $\boldsymbol{A}$ for all three algorithms for a range of values of M, the number of measurements. WF-Poisson demonstrated consistently improved NRMSE compared to GS and WF-Gaussian, with an average improvement of 21.7% / 12.5% compared to GS and an average improvement of 16.4% / 12.5% compared to WF-Gaussian, for real/complex true signals. Tab. 1 shows additional NRMSE statistics regarding across different count levels $\mathbb{E}[|\boldsymbol{a}_i'\boldsymbol{x}|^2]$, where WF-Poisson also showed consistent improvement.

Fig. 5 shows the reconstructed images and the corresponding NRMSE w.r.t. $\boldsymbol{x}_{\text{true-C}}$ for the 3 phase retrieval methods. Compared to GS and WF-Gaussian, WF-Poisson improved NRMSE by 25.4% and 15.3%, respectively.

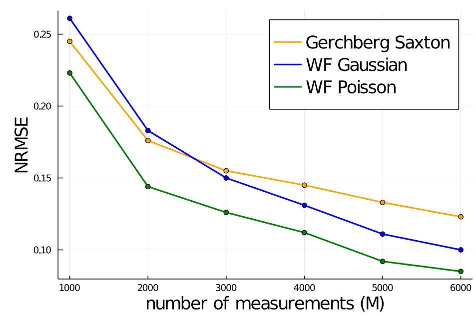


(a) $\boldsymbol{x}_{\text{true-A}}$

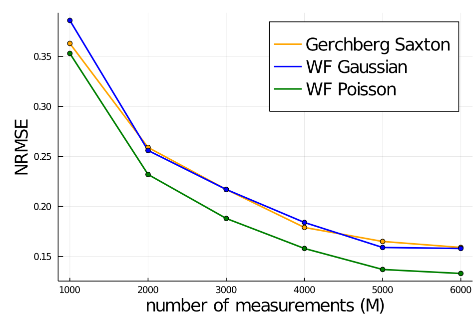


(b) $\boldsymbol{x}_{\text{true-B}}$

Fig. 3. Cost function value versus time. The average of $|\boldsymbol{a}_i'\boldsymbol{x}|^2$ is set to 2 and $M = 3000$.



(a) NRMSE for real case ($\boldsymbol{x}_{\text{true-A}}$)



(b) NRMSE for complex case ($\boldsymbol{x}_{\text{true-B}}$)

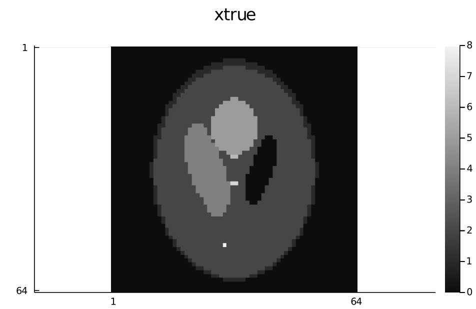
Fig. 4. NRMSE comparison across different number of measurements. The average of $|\boldsymbol{a}_i'\boldsymbol{x}|^2$ is set to 2.

4. DISCUSSION AND CONCLUSION

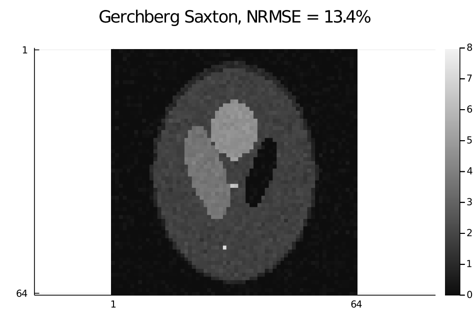
With simulated results showing consistently lower NRMSE tested on real/complex true signals/images, we demonstrated our modified Wirtinger Flow algorithm based on the Poisson ML model has superior accuracy than algorithms based on Gaussian ML model (Wirtinger Flow and Gerchberg Saxton), for measurements having Poisson noise. The WF-Poisson algorithm also demonstrated substantially faster speed in terms of decreasing of NRMSE versus time in this setting, compared to the Gaussian ML algorithms. Future works include investigating algorithms that can also handle non-smooth regularizers (i.e., ℓ_1 norm) and the case when $b_i = 0$, comparing with more algorithms in terms of efficiency and accuracy, and further optimizing algorithms derived from the Poisson noise model.

$\lvert \boldsymbol{a}_i' \boldsymbol{x} \rvert^2 = 3$	Real signal ($\boldsymbol{x}_{\text{true-A}}$)					
NRMSE(%)/M	1000	2000	3000	4000	5000	6000
WF Gaussian	18.6	13.3	11.1	10.0	8.3	8.0
WF Poisson	**16.6**	**11.1**	**9.4**	**8.2**	**6.9**	**6.7**
GS	17.9	13.2	11.4	10.7	9.6	9.2
$\lvert \boldsymbol{a}_i' \boldsymbol{x} \rvert^2 = 3$	Complex signal ($\boldsymbol{x}_{\text{true-B}}$)					
NRMSE(%)/M	1000	2000	3000	4000	5000	6000
WF Gaussian	33.7	20.1	16.3	14.0	13.4	11.8
WF Poisson	**28.4**	**16.8**	**13.8**	**11.8**	**11.1**	**9.8**
GS	30.6	18.9	15.9	13.6	12.8	12.2
$\lvert \boldsymbol{a}_i' \boldsymbol{x} \rvert^2 = 4$	Real signal ($\boldsymbol{x}_{\text{true-A}}$)					
NRMSE(%)/M	1000	2000	3000	4000	5000	6000
WF Gaussian	16.6	11.6	10.3	8.7	7.5	7.1
WF Poisson	**14.6**	**9.5**	**8.4**	**6.7**	**6.2**	**5.6**
GS	15.9	10.5	10.2	8.2	8.1	7.4
$\lvert \boldsymbol{a}_i' \boldsymbol{x} \rvert^2 = 4$	Complex signal ($\boldsymbol{x}_{\text{true-B}}$)					
NRMSE(%)/M	1000	2000	3000	4000	5000	6000
WF Gaussian	23.7	16.6	14.3	13.0	10.3	9.9
WF Poisson	**21.2**	**14.4**	**11.9**	**10.1**	**8.5**	**7.9**
GS	22.6	16.1	13.4	11.3	10.1	9.4
$\lvert \boldsymbol{a}_i' \boldsymbol{x} \rvert^2 = 5$	Real signal ($\boldsymbol{x}_{\text{true-A}}$)					
NRMSE(%)/M	1000	2000	3000	4000	5000	6000
WF Gaussian	14.8	9.7	8.8	8.2	7.1	6.2
WF Poisson	**13.2**	**8.3**	**7.3**	**6.8**	**5.5**	**5.0**
GS	14.6	9.9	8.7	7.8	6.9	6.3
$\lvert \boldsymbol{a}_i' \boldsymbol{x} \rvert^2 = 5$	Complex signal ($\boldsymbol{x}_{\text{true-B}}$)					
NRMSE(%)/M	1000	2000	3000	4000	5000	6000
WF Gaussian	22.5	15.5	12.8	11.0	9.4	8.5
WF Poisson	**18.7**	**13.0**	**10.5**	**8.9**	**7.6**	**7.2**
GS	19.6	14.5	12.0	10.0	8.6	8.5

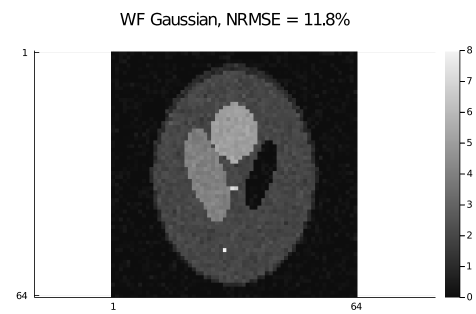
Table 1. NRMSE results comparing the 3 phase retrieval methods (WF-Gaussian, WF-Poisson and GS) for various number of measurements M and mean Poisson count levels. The WF method designed for Poisson noise consistently has the lowest error for all cases.



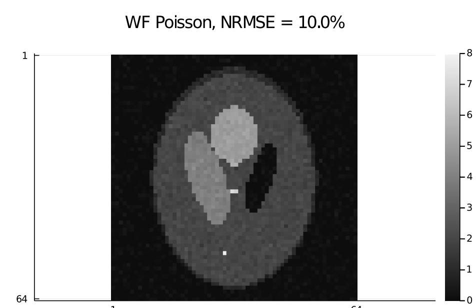
(a) True image ($\boldsymbol{x}_{\text{true-C}}$)



(b) Reconstructed image by GS, NRMSE = 13.4%



(c) Reconstructed image by WF Gaussian, NRMSE = 11.8%



(d) Reconstructed image by WF Poisson, NRMSE = 10.0%

Fig. 5. Reconstructed image and corresponding NRMSE for $\boldsymbol{x}_{\text{true-C}}$ regarding these 3 algorithms. Number of measurements (M) and the average of $\lvert \boldsymbol{a}_i' \boldsymbol{x} \rvert^2$ were set to 2e5 and 2, respectively.

5. REFERENCES

[1] K. Jaganathan, Y. C. Eldar, and B. Hassibi, "Phase retrieval: an overview of recent developments," 2015.

[2] P. Grohs, S. Koppensteiner, and M. Rathmair, "Phase retrieval: uniqueness and stability," *SIAM Review*, vol. 62, no. 2, pp. 301–50, 2020.

[3] P. Jaming, "Phase retrieval techniques for radar ambiguity problems.," *Journal of Fourier Analysis and Applications.*, vol. 5, no. 4, pp. 309–329, 1999.

[4] R. P. Millane, "Phase retrieval in crystallography and optics," *Journal of the Optical Society of America A*, vol. 7, no. 3, pp. 394–411, 1990.

[5] J. Dainty and James Fienup, "Phase retrieval and image reconstruction for astronomy," *Image Recovery: Theory Appl*, vol. 13, 01 1987.

[6] L. R. Rabiner and B.-H. Juang, *Fundamentals of Speech Recognition*, Englewood Cliffs, NJ: Prentice Hall, 1993.

[7] T. Qiu, P. Babu, and D. P. Palomar, "PRIME: phase retrieval via majorization-minimization," *IEEE Trans. Sig. Proc.*, vol. 64, no. 19, pp. 5174–86, Oct. 2016.

[8] E. Candes, X. Li, and M. Soltanolkotabi, "Phase retrieval via Wirtinger flow: Theory and algorithms," *IEEE Transactions on Information Theory*, vol. 61, no. 4, pp. 1985–2007, April 2015.

[9] 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–&, 1972.

[10] P. Thibault and M. Guizar-Sicairos, "Maximum-likelihood refinement for coherent diffractive imaging," *New J. of Phys.*, vol. 14, no. 6, pp. 063004, June 2012.

[11] A. Goy, K. Arthur, S. Li, and G. Barbastathis, "Low photon count phase retrieval using deep learning," *Phys. Rev. Lett.*, vol. 121, no. 24, pp. 243902, Dec. 2018.

[12] R. Xu, M. Soltanolkotabi, J. P. Haldar, W. Unglaub, J. Zusman, A. F. J. Levi, and R. M. Leahy, "Accelerated Wirtinger flow: A fast algorithm for ptychography," 2018.

[13] D. A. Barmherzig and J. Sun, "Low-photon holographic phase retrieval," in *OSA: Computational Optical Sensing and Imaging*, 2020.

[14] 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. Am. A*, vol. 38, no. 1, pp. 71–9, Jan. 2021.

[15] D. L. Snyder, C. W. Helstrom, A. D. Lanterman, M. Faisal, and R. L. White, "Compensation for readout noise in CCD images," *J. Opt. Soc. Am. A*, vol. 12, no. 2, pp. 272–83, Feb. 1995.

[16] K. Choi and A. D. Lanterman, "Phase retrieval from noisy data based on minimization of penalized I-divergence," *J. Opt. Soc. Am. A*, vol. 24, no. 1, pp. 34–49, Jan. 2007.

[17] L. Bian, J. Suo, J. Chung, X. Ou, C. Yang, F. Chen, and Q. Dai, "Fourier ptychographic reconstruction using Poisson maximum likelihood and truncated Wirtinger gradient," *Nature Sci. Rep.*, vol. 6, no. 1, 2016.

[18] E. J. Candes, Y. C. Eldar, T. Strohmer, and V. Voroninski, "Phase retrieval via matrix completion," *SIAM J. Imaging Sci.*, vol. 6, no. 1, pp. 199–225, 2013.

[19] Y. Chen and E. J. Candes, "Solving random quadratic systems of equations is nearly as easy as solving linear systems," *Comm. Pure Appl. Math.*, vol. 70, no. 5, pp. 822–83, May 2017.

[20] H. Chang and S. Marchesini, "Denoising Poisson phaseless measurements via orthogonal dictionary learning," *Optics Express*, vol. 26, no. 16, pp. 19773–96, Aug. 2018.

[21] H. Chang, Y. Lou, Y. Duan, and S. Marchesini, "Total variation–based phase retrieval for Poisson noise removal," *SIAM journal on imaging sciences*, vol. 11, no. 1, pp. 24–55, 2018.

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

[23] I. Kang, F. Zhang, and G. Barbastathis, "Phase extraction neural network (PhENN) with coherent modulation imaging (CMI) for phase retrieval at low photon counts," *Optics Express*, vol. 28, no. 15, pp. 21578–600, July 2020.

[24] M. Makitalo and A. Foi, "Optimal inversion of the generalized Anscombe transformation for Poisson-Gaussian noise," *IEEE Transactions on Image Processing*, vol. 22, no. 1, pp. 91–103, 2013.