2023 IEEE 9th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP) | 979-8-3503-4452-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/CAMSAP58249.2023.10403449

SPARSE OVERDISPERSED PHOTON-LIMITED SIGNAL RECOVERY WITH UPPER AND LOWER BOUNDS

Yu Lu and Roummel F. Marcia

Department of Applied Mathematics, University of California, Merced, Merced, CA, 95343 USA

ABSTRACT

This study addresses the challenge of reconstructing sparse signals, a frequent occurrence in the context of overdispersed photon-limited imaging. While the noise behavior in such imaging settings is typically modeled using a Poisson distribution, the negative binomial distribution is more suitable in overdispersed scenarios where the noise variance exceeds the signal mean. Knowledge of the maximum and minimum signal intensity can be effectively utilized within the computational framework to enhance the accuracy of signal reconstruction. In this paper, we use a gradient-based method for sparse signal recovery that leverages a negative binomial distribution for noise modeling, enforces bound constraints to adhere to upper and lower signal intensity thresholds, and employs a sparsity-promoting regularization term. The numerical experiments we present demonstrate that the incorporation of these features significantly improves the reconstruction of sparse signals from overdispersed measurements.

Index Terms— Negative binomial distribution, low-count imaging, bounded sparse reconstruction, gradient-based optimization.

1. INTRODUCTION

Applications involving low-count measurements are commonplace in a variety of fields ranging from medical imaging [1, 2, 3] and astronomy [4, 5], to network traffic analysis [6, 7] and epidemiology [8, 9]. Specifically in imaging, the reconstruction of discrete, low-photon signals typically employs the Poisson process model [10]. Numerous methods have been developed to address the ensuing reconstruction problem [11, 12, 13, 14]. Nonetheless, the negative binomial distribution serves as a more comprehensive model, encompassing the Poisson distribution as a special case [15, 16, 17].

The negative binomial probability mass function is expressed as

$$P(y|r,p) = \binom{y+r-1}{y}(1-p)^y p^r,$$

which can be interpreted as the probability of observing y successful events within a sequence of independent and identically distributed Bernoulli trials, before the occurrence of the r^{th} failure event, where each trial carries a failure probability of p.

In this context, r is commonly recognized as the dispersion parameter. The mean, μ, is defined as $\mu = r(1-p)/p$ (see [18] for a detailed explanation), which in turn yields the formula $p = r/(r+\mu)$. Note that

$$\lim_{r\to\infty} P(y|r,p) = \frac{e^{-\mu}\mu^y}{y!},$$

which corresponds to the probability mass function of a Poisson-distributed random variable with mean μ.

2. PROBLEM FORMULATION

Within a low photon context, we model the observation y of photons at the detector using the negative binomial assumption: $y \sim \text{NB}(r,p)$. Accordingly, for a vector of observations y, our model can be redefined as

$$y_i \sim \text{NB}\left(r_i, \frac{r_i}{r_i + (Af^*)_i}\right),$$

where f^* is the scene of interest and A is a mapping that linearly projects the scene f^* onto a set of expected measurements Af^*. Thus, the probability of observing the data y is

$$P(y|f) = \prod_{i=1}^{m} \binom{r_i+y_i-1}{y_i}\left(\frac{r_i}{r_i+(Af)_i}\right)^{r_i}\left(\frac{(Af)_i}{r_i+(Af)_i}\right)^{y_i}. \quad (1)$$

Applying the maximum likelihood principle, we aim to maximize the probability of observing the vector y in (1), accomplished by minimizing the corresponding negative binomial log-likelihood function:

$$F(f) \equiv \sum_{i=1}^{m} (r_i+y_i)\log(r_i+(Af)_i) - y_i\log((Af)_i). \quad (2)$$

When additional structural information is available (e.g., sparsity), a regularization term $\text{pen}(f)$ can be incorporated that promotes sparsity in the reconstruction, such as the ℓ_1 norm.

This work is partly supported by National Science Foundation grants DMS 1840265 and IIS 1741490.

(a) f^* (b) y_{Poisson} (c) $y_{\text{r=9}}$ (d) $y_{\text{r=3}}$ (e) Mask

Fig. 1. Example of observation model. (a) The true image f^*. (b) Observed measurement y_{Poisson} drawn from a Poisson distribution. (c) Observed measurement $y_{r=9}$ drawn from a NB distribution with $r = 9$. (d) Another observed measurement $y_{r=3}$ drawn from a negative binomial (NB) distribution with dispersion parameter $r = 3$. (e) Mask containing *a priori* structural intensity bounds.

Furthermore, when the maximum and minimum values of the signal intensity are known in regions of the signal, these structural priors can be integrated as bounds on the reconstructed signal. This information leads to the following bounded negative binomial model [17] for sparse image recovery:

$$\begin{aligned} f^* \;=\; \underset{f\in\mathbb{R}^n}{\arg\min} \quad & \Phi(f) \equiv F(f) + \tau\|f\|_1 \\ \text{subject to} \quad & b_L \le f \le b_U, \end{aligned} \tag{3}$$

where $\tau > 0$ is a regularization parameter that balances the data-fidelity term $F(f)$ with the sparsity-promoting term $\|f\|_1$. Note that this formulation is more specific than the typical non-negativity bound $f \ge 0$. **As such, the optimization problem** (3) **is more difficult than one with only non-negativity constraints.** In this work, we describe a gradient-based optimization approach for solving (3) that includes upper and lower bounds that model signal intensity limits.

3. ALGORITHM

We address (3) by solving a sequence of quadratic subproblems using the second-order Taylor series at the current iterate f^j. This is expressed as follows:

$$\begin{aligned} F^j(f) = \; & F(f^j) + (f - f^j)^\top \nabla F(f^j) \\ & + (f - f^j)^\top \nabla^2 F(f^j)(f - f^j). \end{aligned}$$

To streamline this process, we make certain simplifying assumptions. Firstly, we assume that the dispersion parameters r_i are constant, i.e., $r_i = r$ for a given scalar r, applicable to all i. Second, the derivatives of $F(f)$ are given by the following formulas:

$$\nabla F(f) = \sum_{i=1}^{m} \left(\frac{r + y_i}{r + (Af)_i} - \frac{y_i}{(Af)_i} \right) A^\top e_i$$

$$\nabla^2 F(f) = A^\top \left[\sum_{i=1}^{m} \left(\frac{r + y_i}{(r + (Af)_i)^2} - \frac{y_i}{(Af)_i^2} \right) e_i e_i^\top \right] A,$$

where e_i represents the i^{th} column of the $m \times m$ identity matrix. We utilize the Barzilai-Borwein criteria [19] to approximate the second derivative of $F(f)$ as a scalar multiple of the identity matrix, i.e., $\nabla^2 F(f) \approx \alpha_j I$, with $\alpha_j > 0$ as a particular scalar. Thirdly, we presume that f exhibits sparsity or compressibility in some orthonormal basis W, meaning that $f = W\theta$ for a certain sparse vector θ.

Letting $q^j = f^j - \frac{1}{\alpha_j}\nabla F(f^j)$, the quadratic subproblems and corresonding iterates are then given by

$$\begin{aligned} \theta^{j+1} \;=\; \underset{\theta\in\mathbb{R}^n}{\arg\min} \quad & \frac{1}{2}\|W\theta - q^j\|_2^2 + \frac{\tau}{\alpha_j}\|\theta\|_1 \\ \text{subject to} \quad & b_L \le W\theta \le b_U \\ f^{j+1} \;=\; & W\theta^{j+1}. \end{aligned} \tag{4}$$

Following [20], this optimization problem can be solved using the corresponding Lagrangian dual problem, which we describe next.

To apply gradient-based techinques for solving (4), we introduce vectors $u, v \in \mathbb{R}^n$ such that $\theta = u - v$ with $u, v \ge 0$ so that the non-differentiable ℓ_1 norm is now differentiable, i.e., $\|\theta\|_1 = \mathbb{1}^\top(u + v)$. Therefore, (4) becomes

$$\begin{aligned} (u^{j+1}, v^{j+1}) \;=\; \underset{u,v\in\mathbb{R}^n}{\arg\min} \quad & \frac{1}{2}\|u - v - q^j\|_2^2 + \frac{\tau}{\alpha_j}\mathbb{1}^\top(u + v) \\ \text{subject to} \quad & 0 \le u, v \\ & 0 \le W(u - v) - b_L \\ & 0 \le b_U - W(u - v). \end{aligned} \tag{5}$$

The Lagrangian function corresponding to (5) is given by

$$\begin{aligned} \mathcal{L}(u,v,\lambda_1,\lambda_2,\lambda_3,\lambda_4) = \; & \frac{1}{2}\|u - v - q^j\|_2^2 \\ & + \frac{\tau}{\alpha_j}\mathbb{1}^\top(u + v) - \lambda_1^\top u - \lambda_2^\top v \\ & - \lambda_3^\top(W(u - v) - b_L) - \lambda_4^\top(b_U - W(u - v)), \end{aligned} \tag{6}$$

where $\lambda_1, \lambda_2, \lambda_3, \lambda_4 \in \mathbb{R}_+^n$ are the Lagrange multipliers. Differentiating (6) with respect to u and v and setting both deriva-

tives to zero yields the Lagrangian dual function

$$\mathcal{G}(\lambda_1,\lambda_3,\lambda_4) = -\frac{1}{2}\|q^j + \lambda_1 - \frac{\tau}{\alpha_j}\mathbb{1} + W^\top(\lambda_3-\lambda_4)\|_2^2 + \lambda_3^\top b_L - \lambda_4^\top b_U + \frac{1}{2}\|q^j\|_2^2, \quad (7)$$

which is independent of the variables u, v, and λ_2. Letting $\gamma = \lambda_1 - \frac{\tau}{\alpha_j}$, we obtain the Lagrange dual problem

$$\begin{aligned} \underset{\gamma,\lambda_3,\lambda_4\in\mathbb{R}^n}{\text{minimize}} \quad & \frac{1}{2}\|q^j + \gamma + W^\top(\lambda_3-\lambda_4)\|_2^2 - \lambda_3^\top b_L + \lambda_4^\top b_U \\ \text{subject to} \quad & -\frac{\tau}{\alpha_j}\mathbb{1} \le \gamma \le \frac{\tau}{\alpha_j}\mathbb{1}, \quad 0 \le \lambda_3, \lambda_4, \end{aligned} \quad (8)$$

which can be solved by minimizing over $\gamma, \lambda_3, \lambda_4$ alternatingly (see [20] for details).

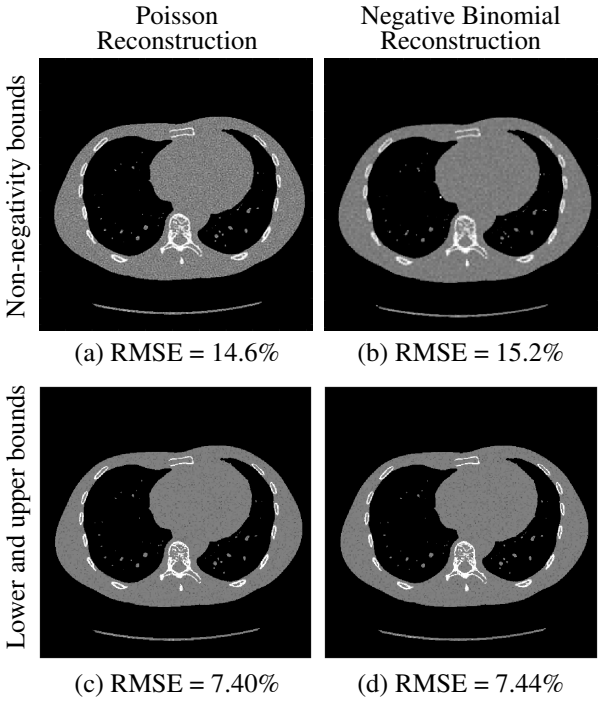
4. EXPERIMENTS

We carried out three separate experiments featuring different noise level realizations. The performance of both the Poisson and negative binomial models was analyzed in light of simple non-negativity constraints as well as general lower and upper bound constraints. The specifics of the experiments are as follows:

- **Experiment I:** 2D data drawn from a Poisson distribution.
- **Experiment II:** 2D data drawn from a negative binomial distribution with dispersion parameter $r = 9$.
- **Experiment III:** 2D data drawn from a negative binomial distribution with dispersion parameter $r = 3$.

In all three experiments, we reconstructed a blurry and noisy image (from the Matlab Medical Imaging Toolbox) of dimensions 512×512. Each experiment deal with distinct noise levels: the noise in the first experiment is drawn from a Poisson distribution, the second from a negative binomial distribution with $r = 3$, and the third from a negative binomial distribution with $r = 9$. For each experiment, we set $b_L = 0$ and $b_U = \max_{i,j}\{f^*_{i,j}\}$. **Every experiment was conducted ten times with ten different noise realizations**, and the average results are presented.

We compared our proposed method to the Poisson model with non-negativity constraints [14], the Poisson model with general lower and upper bounds [20], and the negative binomial model with non-negativity constraints [21]. The percentage root-mean-square error ($\text{RMSE}(\%) = 100 \cdot \|\hat{f} - f^*\|_2/\|f^*\|_2$) served as a measure of the distance between the computed solution $\hat{f}$ and the true solution f^*. Each experiment was limited to a maximum of 100 iterations, as no significant improvements were observed beyond this point. For the negative binomial models (both with non-negativity constraints and general lower and upper bounds), the exact value of r was used. However, this parameter can be estimated using method-of-moment methods [22], maximum quasi-likelihood methods [23], or cross-validation [24]. In our experiments, to avoid error bias from estimating r, we used the exact value of the parameter.



(a) RMSE = 14.6% (b) RMSE = 15.2%

(c) RMSE = 7.40% (d) RMSE = 7.44%

Fig. 2. **Results for Experiment I:** 2D data drawn from a Poisson distribution. (a) The Poisson reconstruction with non-negativity bounds (RMSE = 14.6%). (b) The negative binomial (NB) reconstruction with non-negativity bounds (RMSE = 15.2%). (c) The Poisson reconstruction with lower and upper bounds (RMSE = 7.40%). (d) The NB reconstruction with lower and upper bounds (RMSE = 7.44%).

4.1. Experiment I (Poisson): 2D data drawn from a Poisson distribution

Notably, a Poisson distribution can also be interpreted as a negative binomial distribution when r is significantly large ($r \gg 1$). In these experiments, we assigned a dispersion parameter value of $r = 10^3$. Due to the fact that the r in our experiment is not infinite, the Poisson model slightly outperforms the negative binomial model. However, this difference diminishes as r increases. Interestingly, when bounds are incorporated within the maximum likelihood approach, the outcomes between the Poisson and negative binomial models become fairly comparable.

A key observation in this experiment was that the largest gaps in reconstruction results emerged between methods using non-negativity constraints and those employing lower and upper bounds. In other words, the advantage gained from integrating lower and upper bounds into the reconstructions lessened as the measurements grew noisier.

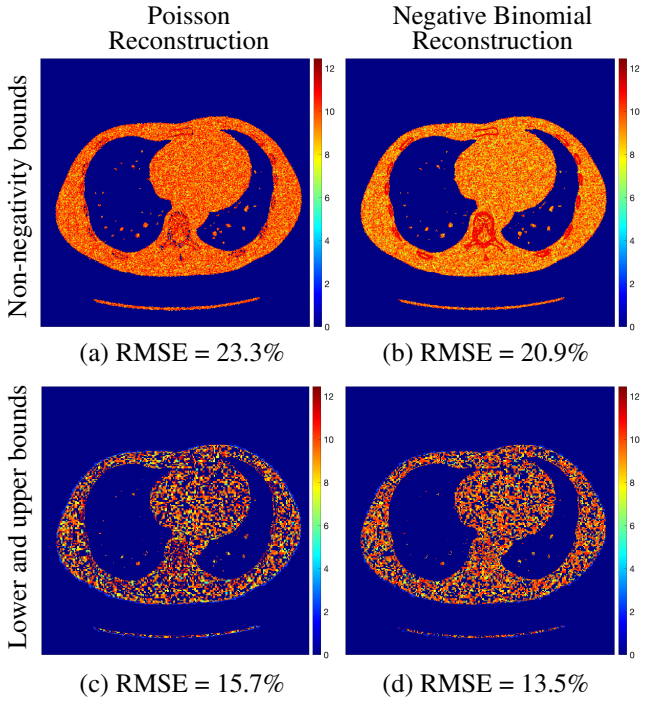


Fig. 3. **Results for Experiment II: 2D data drawn from a negative binomial distribution with $r = 9$.** Here, we present the log error plot of the difference between ground truth f^* and the reconstruction $\hat{f}$ (given component-wise by $\log(|\hat{f} - f^*| + 1)$). (a) The Poisson reconstruction error using non-negativity constraints (RMSE = 23.3%). (b) The negative binomial (NB) reconstruction error using non-negativity constraints (RMSE = 20.9%). (c) The Poisson reconstruction error using lower and upper bounds (RMSE = 15.7%). (d) The NB reconstruction error using lower and upper bounds (RMSE = 13.5%.)

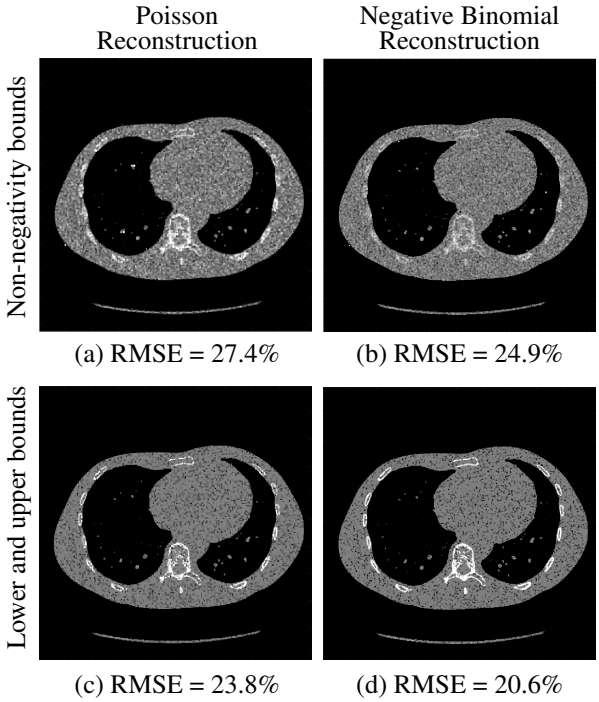


Fig. 4. **Results of Experiment III: 2D data drawn from a negative binomial distribution with $r = 3$.** (a) The Poisson reconstruction with non-negativity bounds (RMSE = 27.4%). (b) The negative binomial (NB) reconstruction with non-negativity bounds (RMSE = 24.9%). (c) The Poisson reconstruction with lower and upper bounds (RMSE = 23.8%). (d) The NB reconstruction with lower and upper bounds (RMSE = 20.6%).

4.2. Experiment II: 2D data drawn from a negative binomial distribution with dispersion parameter $r = 9$

This experiment mirrors Experiment I, with the key difference being that the observations here are drawn from a negative binomial distribution with a marginally smaller dispersion parameter, $r = 9$. Due to the noisier measurement data compared to Experiment I, we anticipate higher RMSEs for all four models. We underscore the significant disparity and improvement gained from implementing lower and upper bounds over just non-negativity constraints, by presenting the log of the error in the reconstruction.

4.3. Experiment III: 2D data drawn from a negative binomial distribution with dispersion parameter $r = 3$

In this third experiment, we employ the observation that are drawn from a negative binomial distribution with a dispersion parameter $r = 3$. The measurements for this experiment are the noisiest of all. We note that the negative binomial reconstructions yield smaller RMSE's than Poisson reconstructions. Significant improvements are observed when we use a negative binomial model and, even more prominently, when we incorporate bounds on the reconstructions (see Fig. 4).

5. CONCLUSION

Low-count signal reconstruction is prevalent in many real-world applications. In this work, we integrate general lower and upper bounds into our maximum likelihood methodology to reconstruct low-photon count signals. Specifically, we focus on the negative binomial model with an orthonormal basis penalty and propose a sequential quadratic optimization method to solve the sparse negative binomial log-likelihood problem. Through a series of three numerical experiments, we tested four models, with the measurements drawn from negative binomial and Poisson distributions. Our results show that the negative binomial model and the incorporation of lower and upper bounds on the reconstructions can significantly improve the accuracy of the reconstructions. This improvement becomes more pronounced as the level of noise in the measurements increases.

6. REFERENCES

[1] J. A. Seibert and J. M. Boone, "X-ray imaging physics for nuclear medicine technologists. Part 2: X-ray interactions and image formation," *Journal of Nuclear Medicine Technology*, vol. 33, no. 1, pp. 3–18, 2005.

[2] H. Su, F. Xing, and L. Yang, "Robust cell detection of histopathological brain tumor images using sparse reconstruction and adaptive dictionary selection," *IEEE Transactions on Medical Imaging*, vol. 35, no. 6, pp. 1575–1586, 2016.

[3] X. Ye, Y. Chen, and F. Huang, "Computational acceleration for MR image reconstruction in partially parallel imaging," *IEEE Transactions on Medical Imaging*, vol. 30, no. 5, pp. 1055–1063, 2010.

[4] R. L. Lucke and L. J. Rickard, "Photon-limited synthetic-aperture imaging for planet surface studies," *Applied optics*, vol. 41, no. 24, pp. 5084–5095, 2002.

[5] S. Cabanillas, M. Guillaume, and P. Réfrégier, "Reconstruction of astronomical images from low-photon image sequence: study for Gaussian displacements," in *IWISPA 2000. Proceedings of the First International Workshop on Image and Signal Processing and Analysis*. IEEE, 2000, pp. 35–40.

[6] S. B. Slimane and T. Le-Ngoc, "A doubly stochastic Poisson model for self-similar traffic," in *Proceedings IEEE International Conference on Communications ICC '95*, 1995, vol. 1, pp. 456–460 vol.1.

[7] E. del Arco, E. Morgado, M. I. Chidean, J. Ramiro-Bargueno, I. Mora-Jiménez, and A. J. Caamano, "Sparse vehicular sensor networks for traffic dynamics reconstruction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 5, pp. 2826–2837, 2015.

[8] E. L. Frome and H. Checkoway, "Use of Poisson regression models in estimating incidence rates and ratios," *American Journal of Epidemiology*, vol. 121, no. 2, pp. 309–323, 1985.

[9] G. Zou, "A modified Poisson regression approach to prospective studies with binary data," *American Journal of Epidemiology*, vol. 159, no. 7, pp. 702–706, 2004.

[10] D. L. Snyder, *Random point processes*, Wiley-Interscience, 1975.

[11] J. A. Fessler and A. O. Hero, "Penalized maximum-likelihood image reconstruction using space-alternating generalized EM algorithms," *IEEE Transactions on Image Processing*, vol. 4, no. 10, pp. 1417–1429, 1995.

[12] B. Zhang, J. M. Fadili, and J.-L. Starck, "Wavelets, ridgelets, and curvelets for Poisson noise removal," *IEEE Transactions on image processing*, vol. 17, no. 7, pp. 1093–1108, 2008.

[13] M. A. T. Figueiredo and J. M. Bioucas-Dias, "Restoration of Poissonian images using alternating direction optimization," *IEEE Transactions on Image Processing*, vol. 19, no. 12, pp. 3133–3145, 2010.

[14] Z. T. Harmany, R. F. Marcia, and R. M. Willett, "This is SPIRAL-TAP: Sparse Poisson Intensity Reconstruction ALgorithms:Theory and Practice," *IEEE Transactions on Image Processing*, vol. 21, no. 3, pp. 1084–1096, 2012.

[15] C. G. Taborda, F. Pérez-Cruz, and D. Guo, "New information-estimation results for Poisson, binomial and negative binomial models," in *2014 IEEE International Symposium on Information Theory*. IEEE, 2014, pp. 2207–2211.

[16] V. Santarelli, M. F. Positano and L. Landini, "Measured PET data characterization with the negative binomial distribution model," *Journal of Medical and Biological Engineering*, vol. 37, pp. 299–312, 2017.

[17] D. Guo, "On information-estimation relationships over binomial and negative binomial models," in *2013 IEEE International Symposium on Information Theory*. IEEE, 2013, pp. 459–463.

[18] M. H. DeGroot and M. J. Schervish, *Probability and statistics*, Pearson Education, 2012.

[19] J. Barzilai and J. M. Borwein, "Two-Point Step Size Gradient Methods," *IMA Journal of Numerical Analysis*, vol. 8, no. 1, pp. 141–148, 01 1988.

[20] L. Adhikari and R. F. Marcia, "Bounded sparse photon-limited image recovery," in *2016 IEEE International Conference on Image Processing*. IEEE, 2016, pp. 3508–3512.

[21] Y. Lu and R. F. Marcia, "Negative binomial optimizatin for low-count overdispersed sparse signal reconstruction," in *Accpeted*, 2023.

[22] S. J. Clark and J. N. Perry, "Estimation of the negative binomial parameter κ by maximum quasi-likelihood," *Biometrics*, pp. 309–316, 1989.

[23] W. W. Piegorsch, "Maximum likelihood estimation for the negative binomial dispersion parameter," *Biometrics*, pp. 863–867, 1990.

[24] C. Gu, "Cross-validating non-Gaussian data," *Journal of Computational and Graphical Statistics*, vol. 1, no. 2, pp. 169–179, 1992.